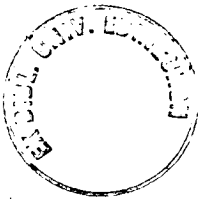
Studies on structure/function relationship in the σ^{70} subunit of RNA polymerase and on *scrP*, an erstwhile candidate minor sigma gene of *Escherichia coli*.

Taciana Kaściukovič

A thesis presented for the Degree of Ph.D.

Institute of Cell and Molecular Biology
The University of Edinburgh

1999



маім маці і бацьку

Table of Contents

Declaration vii

Acknowledgements viii

Abstract ix

Abbreviations x

1. INTRODUCTION 1

1.1. Sigma subunit of RNA polymerase. 2

1.1.1. Transcription initiation by bacterial RNA polymerase 2

1.1.2. Conservation among members of σ^{70} family. 6

1.1.3. Amino acids in σ subunits involved in promoter recognition and opening. 8

1.1.4. Possible roles of region 3 in transcription initiation 13

1.1.5. Use of cysteine residues and modification producing free radicals in structural and functional studies of σ^{70}. 14

1.1.6. The region of σ^{70} between the conserved regions 1.2 and 2.1. 16

1.2. *scrP*, an erstwhile candidate minor sigma gene of *E. coli*. 19

1.2.1. Discovery and early studies of ScrP. 19

1.2.2. Characterisation of *scrP* by gene disruption 22

1.2.3. ScrP protein. 23

1.2.4. Polypeptides with sequence similarities to ScrP 24

1.2.5. Environment of the *scrP* gene on the bacterial chromosome 26

1.2.6. The *mtgA* gene and its product 27

1.2.7. The arcB gene and ArcB/ArcA signal transduction system 30
Two-component regulatory systems. 30
Multi-step phosphorelay regulatory systems. 33
The ArcB/ArcA signal transduction system 34
The *arcB* gene and its product 35
Possible signals for Arc system activation. 38
The *arcA* gene and its product. 40
Phosphorelay within the Arc system. 41
Autophosphorylation of ArcB 42
Phosphorylation of ArcA *in vitro*. 43
Signal decay in the ArcB/ArcA phosphorelay system 44
SixA as a specific phosphatase for ArcB. 46
Does "cross-talk" between members of non-cognate phosphorelay families occur under physiological conditions? 46

2. MATERIALS AND METHODS 48

2.1. Bacterial strains, phages, plasmids, oligonucleotides and proteins 49

2.2. Growth media and buffers 57
2.2.1. Growth media. 57
2.2.2. Commonly used buffers 58
2.2.3. Minimal medium supplements 59
2.2.4. Growth of bacteria. 60
2.2.5. Selection of antibiotic resistance. 61
2.2.6. Dye sensitivity tests. 61

2.3. Bacterial and phage techniques 62
2.3.1. Preparation of competent cells and transformation with plasmid DNA using DMSO. 62
2.3.2. Preparation of competent cells and transformation with plasmid DNA using calcium chloride 63
2.3.3. Frozen storage of bacterial strains. 63
2.3.4. Preparation of phage P1 plate lysates. 64
2.3.5. Phage P1-mediated transduction 64
2.3.6. Phage P1-mediated transfer of plasmid DNA to chromosome. 65
2.3.7. Purification of minicells for in vivo protein synthesis studies. 65
2.3.8. Labelling of plasmid-encoded proteins in minicells. 66
2.3.9. Labelling of plasmid-encoded proteins in maxicells. 67

2.4. DNA Techniques 67
2.4.1. Plasmid preparation. 67
2.4.2. Phenol extraction and ethanol precipitation of DNA. 69
2.4.3. Determination of DNA concentrations. 70
2.4.4. Restriction of DNA. 70
2.4.5. Ligation of DNA 71
2.4.6. 'Filling in' of recessed 3' termini. 71
2.4.7. Agarose gel electrophoresis. 72
2.4.8. Isolation of DNA from agarose gel slices. 72
2.4.9. Labelling DNA fragments by random priming. 73
2.4.10. Preparation of chromosomal DNA. 73
2.4.11. Preparation of bacteriophage M13 DNA. 74
2.4.12. Southern blotting procedures 75
2.4.13. Amplification of DNA by polymerase chain reaction 77
2.4.14. Site-directed mutagenesis 78
2.4.15. DNA sequencing 80

2.5. Protein Techniques 80
2.5.1. Polyacrylamide gel electrophoresis of proteins. 80
2.5.2. Western blotting procedures 84
2.5.3. Overexpression of proteins in *E. coli* cells. 86
2.5.4. Isolation of inclusion bodies 86
2.5.5. Protein purification from inclusion bodies 87
2.5.6. Column chromatography. 88
2.5.7. Protein concentration determination. 89
2.5.8. Preparation of total and outer E. coli membranes for protein analysis. 89
2.5.9. Preparation of periplasmic protein fraction. 90
2.5.10. Hybrid σ assay. 90
2.5.11. β-galactosidase activity assay. 91
2.5.12. Selection for kat ($RpoS^-$) phenotype. 92

2.6. RNA polymerase techniques 93
2.6.1. Promoter DNA templates for transcription. 93
2.6.2. *In vitro* transcription. 93
2.6.3. Gel-shift assay. 95

2.7. Sequence analysis 95

2.8. Image analysis and densitometry 96

3. STUDIES ON STRUCTURE/FUNCTION RELATIONSHIP IN THE SIGMA-70 SUBUNIT OF RNA POLYMERASE 98

3.1. Substitution of single *rpoD* codons for amino acid residues in regions 2.3, 2.4, 3.1 and 3.2 of σ^{70} with cysteine 99

3.2. The *in vivo* activities of sigma subunits with single cysteine replacements 102

3.3. Hydrophobicity of $\sigma^{70\Delta 245\text{-HMK}}$ and its derivatives with single cysteine replacements 103

3.4. Reversibility of oligomerisation of $\sigma^{70\Delta 245\text{-HMK}}$ and of its single cysteine replacement derivatives. 108

3.5. Activity of $E\sigma^{70\Delta 245\text{-HMK}}$ and the derivatives *in vitro*. 111

3.6. T7A1 DNA promoter binding and open complex formation 111

3.7. The activity of $E\sigma^{70\Delta 245\text{-HMK}}$ and the derivatives from T7A1 and λP_R promoters *in vitro*. 114

3.8. Abortive synthesis from λP_R promoter 115

3.9. Sensitivity of both productive and abortive complexes to heparin. 120

Discussion 121

4. *SCRP* GENE AND ITS PRODUCT 128

4.1. Theoretical analysis of ScrP and MtgA polypeptide sequences 129

4.2. *In vivo* expression of the plasmid-borne genes of the *arcB* operon. 133

4.2.1. Expression of *scrP* and *mtgA* in mini- and maxicells 133

4.2.2. Overexpression of *scrP* and *mtgA* using a T7 RNA polymerase-driven promoter 142

4.2.3. Cell fractionation. 148

4.3. Analysis of expression of the genes in the *arcB* operon. 151

4.3.1. Cloning of the *Eco*R I–*Bam*H I fragment of pACE1 into pUC18 151

4.3.2. *lacZ* fusion construction 152

4.3.3. Replacement of the chromosomal copies of *arcB*, *scrP* and *mtgA* with their *lacZ* fusions. 155
4.3.4. Confirmation of the structure of TA6, TS6 and TM6. 157
4.3.5. Sensitivity to dyes....159
4.3.6. Complementation tests....160
4.3.7. The expression levels of the *arcB'–'lacZ*, *scrP'–'lacZ* and *mtgA'–'lacZ* fusions under various growth conditions. 162
4.3.8. *scrP* expression appears to be independent of *rpoS*....170
4.3.9. *arcB* and *scrP* expression is independent of *arcA*....171
4.3.10. Expression of the *scrP'–'lacZ* fusion in an arcB background....174

Discussion 179

REFERENCES 182

Appendix

Declaration

I declare that the composition of this thesis was all my own work, except where stated.

Taciana Kaściukovič

December 1999

Acknowledgements

I would like to thank my supervisor Richard Hayward for his guidance and concern, the Darwin trust of Edinburgh and Ken Murray for providing the studentship without which my study in Edinburgh would not be possible, and Nobuo Shimamoto for his guidance and support in his laboratory in National Institute of Genetics in Mishima.

I also wish to acknowledge all members of staff in ICMB, Edinburgh, and in National Institute of Genetics in Japan who have helped me, with special reference to Helen Williamson, Dave Smillie, Jean Beggs, Maurice Gallagher, Hiroki Nagai, Ranjan Sen, and to all in the Media Room.

I would also like to thank Neil McLennan for his support in many ways, my family and all my friends in Edinburgh, Finchley, Mishima and at home for all the good moments they shared with me.

Thank you all.

Abstract

Sigma–70 is the subunit of *E. coli* RNA polymerase which confers specificity for initiation of transcription at most promoters in actively growing cells. The protein shares four conserved regions with major and many alternative sigmas of all studied prokaryotes. However, a stretch of 245 amino acids, lying between conserved regions 1 and 2, is absent from all minor sigmas and even from the similar and functionally homologous sigma-43 of *Bacillus subtilis*. Deletion of these 245 residues from sigma-70 generates a protein which retains significant function *in vitro* and partial *in vivo* function. The three natural cysteine residues of sigma-70 are confined to the deleted region. Plasmids allowing overproduction of the internally deleted protein and of a number of single-cysteine substituted mutant derivatives were constructed, and these were tested for function *in vivo* using a chimaeric sigma/promoter system. The overproduced proteins were isolated employing a novel purification method, and their activities analysed *in vitro* at λ_{PR} and T7A1 promoters. An extraordinary tendency towards reversible aggregation and hydrophobicity was observed. A possible role of the 245 amino acid domain is discussed.

The *scrP* gene encoding sigma cross-reacting protein was originally identified during a search for alternative sigma factors, but later shown not to be a sigma protein. As part of the effort to elucidate its possible function, the regulation of expression of *scrP* gene was studied. Transcriptional fusions of *lacZ* to *scrP*, and to *arcB* and *mtgA*, the genes adjacent to it, were produced, transferred to the *E. coli* chromosome and utilised to analyse the expression of *scrP*. The results indicate that *scrP* belongs to the same operon as *arcB* and *mtgA*, and is transcribed both from its own promoter and from the promoter upstream from *arcB*. Further analysis was used to demonstrate that ScrP is localised in the periplasm of *Escherichia coli*.

Abbreviations

aa	Amino acid residues
Amp	Ampicillin
ATP	Adenosine 5'-triphosphate disodium salt
A^{wv}	Absorbance at wavelength <wv>, nm
bp	Base pair
CAT	Chloramphenicol acetyltransferase
BSA	Bovine serum albumin
Cm	Chloramphenicol
CTP	Cytidine 5'-triphosphate sodium salt
dATP	Deoxy ATP
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleoside 5'-triphosphate
dsDNA	Double-stranded DNA
E	DNA-dependent RNA polymerase, core enzyme
E·σ	DNA-dependent RNA polymerase, holoenzyme
EDTA	Ethylenediaminetetraacetic acid tetrasodium salt
GTP	Guanosine 5'-triphosphate sodium salt
HEPES	*N*-2-Hydroxyethylpiperazine-*N'*-2-ethanesulphonic acid
IPTG	Isopropylthio-β-D-galactoside
Kan	Kanamycin
kb	Kilobase
kDa	Kilodalton
MOPS	3-*N*-Morpholinopropanesulphonic acid
nt	Nucleotide
NTP	Nucleoside 5'-triphosphate
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
Pi	5'-phosphorylated
PIPES	Piperazine N,N'bis (2-ethane sulphonic acid)
PMSF	Phenylmethylsulphonyl fluoride
rbs	Ribosome binding site
RNA	Ribonucleic acid
RNAP	RNA polymerase
rpm	Revolutions per minute

SDS	Sodium dodecyl sulphate
Spc	Spectinomycin
ssDNA	Single-stranded DNA
Str	Streptomycin
TEMED	N,N,N',N',-tetramethylethylethylenediamine
Tet	Tetracycline
Tris	Tris(hydroxymethyl)aminomethane
ts	Temperature sensitive
UTP	Uridine 5'-triphosphate trisodium salt
X-gal	5-Bromo-4-chloro3-indolyl-β-galactoside
X^R	Resistance to antibiotic X
X^S	Sensitive to antibiotic X

The numbering of amino acids in ArcB is according to Iuchi *et al.* [Iuchi, 1990 #50].

Chapter 1

Introduction

1.1. σ subunit of RNA polymerase.

1.1.1. Transcription initiation by bacterial RNA polymerase.

In eubacteria transcription is carried out by DNA dependent RNA polymerase (RNAP). The activities of this enzyme can be described in terms of two functional forms, one of which is the core enzyme involved in RNA polymerisation and the other the holoenzyme responsible for recognition and binding of specific promoter sequences and for initiation of transcription. The core RNAP interacts with a σ subunit, and together these constitute the RNA polymerase holoenzyme in which the σ subunit determines its transcriptional specificity [Burgess *et al.*, 1969; Gross *et al.*, 1992].

The core enzyme (E) involved in RNA synthesis is zinc containing protein which has a minimal composition of four subunits $\beta\beta'\alpha_2$. The individual subunits do not show biological activity, but can self-assemble in a defined order, as was shown *in vitro* and *in vivo* by Yura and Ishihama [1979]. The core enzyme is unable to initiate transcription specifically, but is capable of ribonucleotide polymerisation and termination at some sites. A functional sigma subunit is absolutely required to allow specific transcription initiation by RNA polymerase [Burgess *et al.*, 1969]. The initiating RNA polymerase holoenzyme (Eσ) of *Escherichia coli* has a molecular weight of about 4.5 x 10^5 and contains five subunits:

$$\alpha+\alpha \longrightarrow \alpha_2 \longrightarrow \alpha_2\beta \longrightarrow \underset{E}{\alpha_2\beta\beta'} \longrightarrow \underset{E\sigma}{\alpha_2\beta\beta'\sigma}$$

The four subunits of the core enzyme have molecular sizes 36.5 kDa (α), 150.6 kDa (β) and 155.2 kDa (β′), and are encoded by *rpoA* (α), *rpoB* (β) and *rpoC* (β′) genes [rev. by Smillie, 1994].

The σ subunit is bound relatively weakly to the rest of the enzyme (the core polymerase); it is responsible for specific promoter recognition by RNA polymerase, but is released during the initiation process. The core polymerase then continues to catalyse

phosphodiester bond formation between the growing RNA chain and the next NTP, the identity of which is specified by the sequence of the template strand. After chain termination and dissociation of the ternary complex of the core RNA polymerase, RNA, and DNA, the released core can again bind a σ subunit and start a new cycle of RNA synthesis.

The *Escherichia coli* genome encodes seven σ factors, of which σ^{70} (named for its molecular weight of 70,000) is the most abundant during exponential cell growth (alternative names are given in brackets):

σ^{70} (RpoD)	[Helmann and Chamberlin, 1988]
σ^{32} (RpoH)	*ibid*
σ^{F} (σ^{28})	*ibid*
σ^{E} (RpoE)	*ibid*
σ^{38} (RpoS)	*ibid*
σ^{FecI}	[Lonetto *et al.*, 1992]
σ^{N} (σ^{54})	[Merrick *et al.*, 1993]

The σ subunit is an important determinant of the sequence-specific recognition of promoter DNA. Through the use of different σ factors, RNA polymerase can be targeted to promoters with distinct sequences. In response to various environmental stimuli different σ factors can preferentially bind to core RNA polymerase to initiate transcription of separate sets of genes which are under their control.

Prokaryotic sigma factors can be classified into two families according to their sequence similarities. Members of the first family are similar to the *E. coli* σ^{70} factor and have four regions of homology which will be described later; they specifically recognise –35 and –10 regions of their cognate promoters. Members of the other family are similar to the *E. coli* σ^{54} factor, have very little similarity to σ^{70} family, and recognise –24 and –12 regions of promoters mostly involved in nitrogen regulation of bacteria [Merrick *et al.*, 1993]. Mechanisms of transcription initiation are

different between these two families. Here only σ^{70} family will be discussed.

The σ^{70} family of transcription factors can be further subdivided into several groups. Bacteria usually have one primary sigma factor which is required for transcription initiation at the majority of promoters in exponentially grown cells. Alternative sigma factors direct transcriptional responses to changing environmental conditions. Although all members of σ^{70} family have significant sequence similarities in four conserved regions, they recognise different sequences at the two promoter elements localised –10 and –35 base pairs upstream of the transcription start sites and thus contribute to the control of initiation of transcription [Helmann and Chamberlin, 1988].

Four important promoter elements can be distinguished: two hexamers centred at or near positions 10 and 35 upstream from the transcription start site (designated by their locations as the –10 and –35 regions), the spacer DNA separating them, and a region between –40 and –60 (the UP element). Various compilations of promoter sequences recognised by $E\sigma^{70}$ have established the importance of the 35 and 10 regions. The consensus sequences of these two regions as read on the nontemplate strand are, respectively, TTGACA and TATAAT. Generally the greater the similarity of the 10 and 35 regions to their consensus sequence, the better the promoter functions *in vitro*, as well as *in vivo*; in this regard, well-conserved base pairs (underlined above) may be of greater importance than the less-conserved ones [Helmann and Chamberlin, 1988]. The ability of RNA polymerase to discriminate between promoters with different sequences at the –10 and –35 regions has allowed the basal rate of transcription of various genes to be set at different levels. A consensus length of 17 bp has been established for the spacer between the 10 and 35 regions. Promoters with such a spacer length have been found to be more active *in vitro*, as well as i*n vivo*, than those with shorter or longer spacers [Helmann and Chamberlin, 1988; Dombroski *et al.*, 1996]. An A+T-rich region between positions 40 and 60 (the UP element) is

recognised in some promoters as an additional important determinant of promoter activity. This region is contacted by the α subunit [Igarashi and Ishihama 1991].

Transcription initiation by RNAP can be described in several steps:

$$E + \sigma \rightleftharpoons E\sigma$$

$$E\sigma + P \rightleftharpoons E\sigma\cdot P_c \rightleftharpoons E\sigma\cdot P_i \rightleftharpoons E\sigma\cdot P_{open} \xrightarrow{NTPs} E\sigma\cdot P_{init} \xrightarrow{-\sigma} E\cdot P_{elong}$$

$$E\sigma\cdot P_{open} \rightleftharpoons E\sigma\cdot P_{abortive} \xrightarrow{NTPs} \text{abortive transcripts}$$

where P is promoter DNA, $E\sigma\cdot P_c$ closed complex, $E\sigma\cdot P_i$ intermediate complexes, $E\sigma\cdot P_o$ open complex, $E\sigma\cdot P_{init}$ initiation complex, $E\cdot P_{elong}$ elongation complex, $E\sigma\cdot P_{abort}$ non-productive complex [Gross *et al.*, 1998].

Initially core enzyme binds to the σ subunit, bringing conformational changes both to the core enzyme and to the σ subunit [Dombroski, 1997; Wilson and Dombroski, 1997; Schaubach and Dombroski, 1999; Callaci *et al.*, 1998; Callaci and Heyduk, 1998; Buckle *et al.*, 1999; Craig *et al.*, 1998; deHaseth *et al.*, 1998; Nagai and Shimamoto, 1997; Polyakov *et al.*, 1995]. It has been now established that σ makes contacts with the core enzyme in many regions along its length, notably 1.1, 2.1, 2.2, 2.4, 3.1, 4.1, and 4.2, according to protein footprinting and genetic studies [Gross *et al.*, 1998; Joo *et al.*, 1997; Joo *et al.*, 1998; Nagai and Shimamoto, 1997]. The holoenzyme, but not the core enzyme or σ alone, is able to specifically recognise and bind promoter sequences within the DNA. The holoenzyme bound to promoter DNA forms a closed complex ($E\sigma P_c$) in which DNA is double-stranded. Then EσP undergoes several isomerisation steps ($E\sigma P_i$), during which promoter DNA melts at the –10 element, the melting spreads to the transcriptional start at +1 and beyond, establishing the open complex $E\sigma P_o$. σ subunit is also needed to precisely locate transcription start site. After addition of NTPs and formation of the first phosphodiester bonds ($E\sigma P_{init}$) RNA polymerase leaves the promoter, soon after σ dissociates from the transcriptional complex, and the elongating complex EP_{elong} continues to synthesise RNA [rev. by Gross *et al.*, 1998].

1.1.2. Conservation among members of σ^{70} family.

Each bacterium has several different sigma subunits responsible for the recognition of different sets of promoters. From extensive studies it has been revealed that all σ subunits fall into two major families, with little similarities between them [Lonetto *et al.*, 1992]. The first, and the largest, σ family has a significant degree of similarity to σ^{70}, the primary σ subunit of *Escherichia coli*, which is responsible for the recognition of the majority of promoters in actively growing cells. The other family of sigmas has similarity to σ^{N}, or σ^{54}, of *E. coli*, and is responsible for transcription initiation of genes involved in nitrogen regulation [Gross *et al.*, 1992].

The σ^{70} family can be further subdivided into primary and alternative σ factors [Gross *et al.*, 1992]. The sequences of many of these are known, and four conserved regions have been identified (Fig. 1.1.1.) [rev. by Helmann and Chamberlin, 1988].

Series of studies have shown that region 2.4 interacts directly with −10 elements, and region 4.2 with −35 elements of promoters [reviewed by Gross *et al.*, 1998]. Region 2.5, between 2.4 and 3.1, was proposed as the site of recognition of the "extended −10" motif of the promoter [Barne *et al.*, 1997; Dombroski, 1997; Bown *et al.*, 1999], which can drive factor-independent transcription at some promoters lacking consensus within −35 element [Kumar *et al.*, 1993].

Free σ^{70} alone or core enzyme does not bind promoter DNA specifically [Wellman and Meares, 1991]. It has been shown that it is region 1.1 of σ^{70} that prevents the polypeptide from binding to DNA in the absence of the core enzyme [Dombroski *et al.*, 1992; Dombroski *et al.*, 1993], negatively affecting binding of region 4 to the −35 element of the promoter [Bowers and Dombroski, 1999].

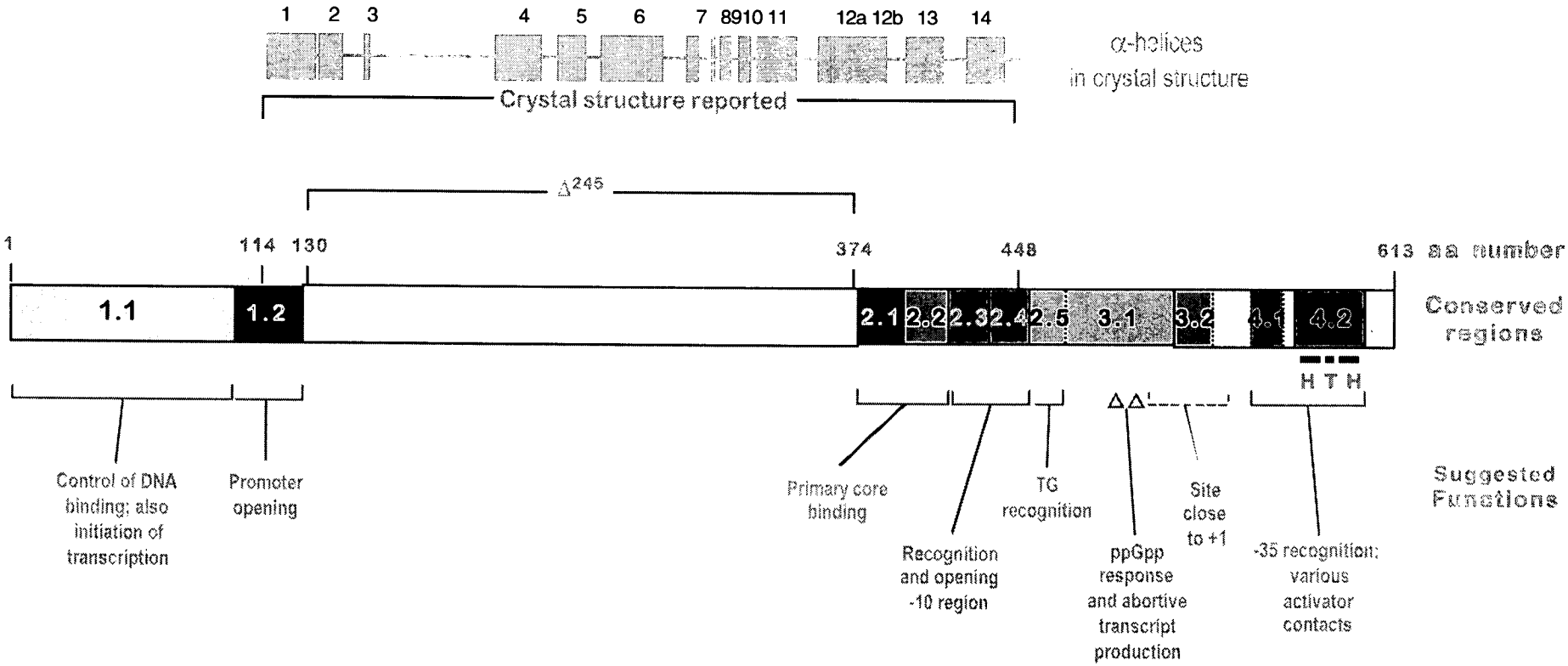


Figure 1.1.1. Primary structure of σ^{70} with conserved regions (1.1–1.4) and some of their known or suggested functions. H T H – predicted helix-turn-helix structure.

However, σ^{70} itself has the ability to recognise and bind the DNA fork between positions –12 and –7 of the transcriptional bubble (a heteroduplex DNA). The identity of the unpaired nucleotide of the nontemplate strand at position –11 is crucial for the recognition [Dombroski, 1997; Guo and Gralla, 1998]. In heteroduplex promoter DNA base recognition by σ^{70} is directed to the nontemplate strand, and G at position –12 or –11 of a consensus promoter was the only base significantly disfavoured by σ^{70} [Roberts and Roberts, 1996].

In the holoenzyme–*lac*UV5 promoter complex sigma cross-links to nucleotides –7 to –2 of exclusively nontemplate strand of the promoter [Simpson, 1979; Siebenlist *et al.*, 1980; Chenchik *et al.*, 1982; Buckle *et al.*, 1999]. Buckle *et al.* observed open complex formation using time-resolved UV irradiation and controlled proteolysis. First distortions at –35 region of the *lac*UV5 promoter occurred, due to intimate contact of σ^{70} with this site, followed by its direct contacts at positions –8, –5 and –3 of the promoter. These contacts occur prior to DNA strand separation. After the fast melting step, and when RNA polymerase already synthesises full-length transcripts, reformation of close contacts of σ^{70} with the –10 region begins, thus reversing the order of σ^{70}–promoter contact events leading to open complex formation [Buckle *et al.*, 1999].

Using UV laser footprinting it was shown that RNA polymerase contacts –35 element of the *malT1* promoter first, and then it spreads the contact in the closed complex to –10, in presence or absence of activator CRP [Eichenberger *et al.*, 1997]. The length of the spacer DNA (17 bp for σ^{70}) between –35 and –10 elements is an important determinant of promoter strength [e.g. Kumar *et al.*, 1995a; Dombroski *et al.*, 1996].

1.1.3. Amino acids in σ subunits involved in promoter recognition and opening.

We were interested in investigating the mechanism of promoter opening. The conserved region 2.3 is known to be involved in DNA melting is [Helmann and Chamberlin, 1988]. Region 2.4 interacts

with the –10 element of promoters. There have been extensive genetic and biochemical studies on amino acid residues thought to be involved in the process of promoter opening. Together with the resolved crystal structure of a fragment of σ^{70} (residues 114–442) [Malhotra *et al.*, 1996] these studies have provided a powerful insight into the mechanisms of DNA recognition and melting.

Many residues in region 2.3 are tolerant to amino acid substitutions. Waldburger and Susskind [1994] used a $\sigma^{70(R588H)}$ which has an altered specificity for the –35 element. They were able to make conservative and non-conservative substitutions at almost all the residues of region 2.3 of $\sigma^{70(R588H)}$, except for T429 (changing this residue to alanine makes the protein inactive) and K426 (only a conservative change to arginine is permissible), while retaining functionality of σ^{70} *in vivo* as measured by the activation of reporter genes from mutated *ant↓–33A* and *lac↓–34A* promoters. Amongst these there were Y430M, single substitutions to cysteine at positions Y421, Y425 and W433, a multiple substitutions containing S428C (S428C Y425F F427L or S428C Y425D F427I) and S428A. Y421C, F427A, W433C and W430A retained only 25–45% of functionality, as judged by their ability to activate the reporter gene.

Helmann's group studied primary sigma σ^{A} of *Bacillus subtilis* with substitutions in region 2.3. They made a number of conserved aromatic amino acid substitutions and found that σ^{A} with Y180A substitution supports transcription initiation; Y184A, Y184L, Y184S, Y189A, Y189F, Y189W, Y189S, Y189L, W192A, W193A, W192S + W193S substitution lead to cold-sensitive transcription from *B. subtilis* P_{ilv} promoter, indicative of melting impairment. Finally, F178A and F186A substitutions lead to the most impaired transcription *in vitro* but even these substitutions can allow transcription from supercoiled templates [Juang and Helmann, 1994] or from a "bubble" template (F178A, Y180A, Y184A, F186A, Y189A, W192A, W193A) [Aiyar *et al.*, 1994]. $E\sigma^{A(Y189A)}$ and $E\sigma^{A(W192A)}$ are defective in nucleation and propagation of the melted region of *B. subtilis* *PtrnS* promoter *in vitro*, while $E\sigma^{A(W193A)}$ can initiate

transcription but is defective in propagation [Juang and Helmann, 1995]. Y180A and Y184A substitutions in σ^A support cell growth in single copy, while overexpressing σ^A with Y189A, W192A and W193A substitutions is lethal to cells [Rong and Helmann, 1994].

$E\sigma^{A(Y180A)}$, $E\sigma^{A(Y189A)}$ and $E\sigma^{A(W192A)}$ are able to bind dsDNA or ssDNA (nontemplate strand) with the consensus −10 element as effectively as $E\sigma^A$. $E\sigma^{A(W193A)}$ shows enhanced, and $E\sigma^{A(Y184A)}$ impaired, binding to ssDNA, but they demonstrate normal binding to double-stranded promoter DNA. $E\sigma^{A(F178A)}$ and $E\sigma^{A(F186A)}$ are grossly defective in DNA binding. Despite good binding of the nontemplate DNA, however, in $E\sigma^{A(Y189A)}$ and $E\sigma^{A(W193A)}$ σ poorly cross-linked to this DNA comparative to the σ in the wild-type holoenzyme, indicating that these amino acids contribute to sequence-selective nontemplate strand binding [Huang *et al.*, 1997]. F178, Y180, Y184, F186, Y189, W192, W193 of σ^{70} correspond to F419, Y421, Y425, F427, Y430, W433 and W444 of σ^A.

Single amino acid substitutions were performed at all positions of region 2.3 and first eight positions of region 2.4 of *B. subtilis* σ^E responsible for endospore development [Jones *et al.*, 1992]. Substitutions at positions 113, 115, 117 and 120 (Y113C, Y113A, S115P, N120A and several C117 substitutions including C117W, C117R) lead to the most severe effects on sporulation. C117 of σ^E aligns with W434 of σ^{70}. $E\sigma^{E(C117R)}$ cannot transcribe from linear templates *in vitro*, but are able to do so from a supercoiled template, with the transcription becoming temperature-dependent, as tested with σ^E-dependent promoter *spoIIID* [Jones and Moran, 1992]. Amongst the substitutions which allow cells to sporulate are P104 to Q or T; I108 to A, S or V; A111 to V; R116 to P or A; L123 to A [Jones *et al.*, 1992]. These amino acids correspond to several of those in *E. coli* σ^{70} that we intended to study (see Fig. 1.1.2). M124T replacement in region 2.4 (aligned with R441 of σ^{70}) suppresses single base pair mutations at the two upstream positions of −10 element of σ^E-dependent *spoIIID*, *spoIID* and *cotEP1* promoters [Tatti *et al.*, 1991; Diederich *et al.*, 1992]. σ^E with double substitution E119 and N120 to isoleucine and glycine (I119 and G120), which are

amino acids that another sigma factor of *B. subtilis*, σ^B, has at these positions, is capable of recognising σ^B-dependent promoter *ctc in vivo*. However, it can no longer transcribe from the σ^E-dependent promoter *spoIID* or from the *ctc* promoter with its first two base pairs GG substituted to CA, the latter being characteristic of σ^E-dependent promoters [Tatti and Moran, 1995]. This suggests that either or both E119 and N120 of σ^E, which align with R436 and Q437 of σ^{70}, directly interact with first two base pairs of σ^E cognate promoters and are responsible for their recognition.

```
σ70      417 DKFEYRRGYKFSTYATWWIRQAITRSIADQARTIRIPVHM 456
σA       176 EKFDYRKGYKFSTYATWWIRQAITRAIADQARTIRIPVHM 215
σE       100 NTFNPEKKIKLATYASRCIENEILMYLRRNNKIRS----- 134
σB        71 KRYDPVVGKSFEAFAIPTIIGEIKRFLRDKTWSVHVPRRI 111
σH        76 RDFKEDKLTSFKAFAELCITRQIITAIKTATRQKHIPLNS 120
             <------2.3------><--------2.4--------->
```

Fig. 1.1.2. Alignment of regions 2.3 and 2.4 σ^{70} of *E. coli* and of σ^A, σ^E, σ^B, and σ^H of *B. subtilis*. Underlined amino acid residues are those substitutions of which are mentioned in the text.

Siegele *et al.* [1989] and Waldburger *et al.* [1990] demonstrated that T440I and Q437H substitutions in σ^{70} region 2.4 suppress T→C mutation at −12 position of σ^{70}-dependent *lac* and *ant* promoters. A438V substitution had no significant effect on the tested mutant promoters [Siegele *et al.*, 1989].

Marr and Roberts [1997] have shown that $E\sigma^{70}$ of *E. coli* binds a template oligonucleotide containing −10 element of $\lambda P_{R'}$ promoter *in vitro*, but does not bind the oligonucleotide with T→C replacement at −12 of this promoter. $E\sigma^{70(Q437H)}$ is able to bind both oligonucleotides specifically.

Dombroski [1997] used fragments of σ^{70} which did not have region 1 and therefore are able to bind duplex promoter DNA, in DNA binding competition assays. She has shown that those fragments which contain region 2 can distinguish the identity of first three nucleotides in the −10 element of the *tac* promoter.

The substitution Q196R in σ^A of *B. subtilis* strongly suppressed a T→C mutation at the upstream position of -10 element of the *spoIIG* promoter, while T199I substitution has only a slight effect (Q196 and T199 of σ^A align with Q437 and T440 of σ^{70}) [Kenney *et al.*, 1989].

Genetic studies on region 2.4 of σ^H, one of *B. subtilis* sporulation sigma factors, have shown that T100I substitution suppresses mutation of -13, and R96A of -12 base pair (the first two positions of the -10 element) of the σ^H-dependent *spoVG* promoter (T100 and R96 of σ^H align with R441 and Q437 of σ^{70}) [Daniels *et al.*, 1990]. $E\sigma^{H(T100I)}$ can also transcribe from the mutant promoter (G→A at position -13 of P_{spoVG}) *in vitro* [Zuber *et al.*, 1989]. I99A substitution causes some impairment of σ^H function *in vivo*, as judged by reduced ability of cells to sporulate, and cells with Q97A substitution in σ^H behaved normally (I99 and Q97 of σ^H align with T440 and A438 of σ^{70}) [Daniels *et al.*, 1990].

W434G substitution in σ^{70} has been shown to cause a 30% decrease in helicity in the free sigma subunit and changes its mobility on SDS–PAGE. The reconstituted holoenzyme displayed decreased level of productive, but normal level of abortive, transcription at T7 A1, *lac*UV5 and *rpsA* promoters. σ^{70} with W433F substitution behaves like wt σ^{70}, whereas W433G substitution leads to a decreased level of transcription *in vitro* [Gopal *et al.*, 1994].

Callaci and Heyduk [1998] replaced tryptophan residues with a fluorescent tryptophan analogue and showed that W433 and W444 become less exposed to a solvent when σ^{70} binds core enzyme. Conformational changes around these residues in region 2.3 also occur after addition of promoter DNA, when these residues become inaccessible to solvent.

Parts of regions 2.3 and 2.4 form helix 14 within the crystal structure of the σ^{70} fragment. This helix is amphipathic, and on the hydrophilic face there are amino acids which are known to interact

with the –10 element of nontemplate strand of the promoter: Y430 and W433 (–10 and –11), Q437 and T440 (–11 and –12) [Malhotra *et al.*, 1996; Gross *et al.*, 1998] (see Fig. 1.1.3.). On the hydrophobic face of the helix there are several absolutely conserved hydrophobic residues in region 2.4 (I435, I439 and I443), which together with other conserved hydrophobic residues on helices 1, 12 and 13, belonging to regions 1.2, 2.1, 2.2 and 2.3, contribute the to hydrophobic core of the σ protein. This amphipathic helix is thought to play a role in melting by acting as a single-strand nucleic acid binding protein and neutralising the DNA phosphate backbone [Malhotra *et al.*, 1996].

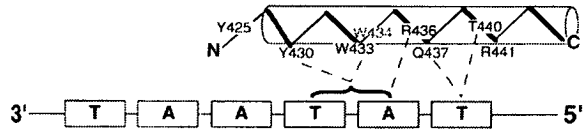


Figure 1.1.3. A proposed interaction between amino acids in helix 14 [Malhotra *et al.*, 1996; Gross *et al.*, 1998] (parts of regions 2.3 and 2.4) with nucleotides of the nontemplate strand of a promoter DNA.

1.1.4. Possible roles of region 3 in transcription initiation.

Region 3 is not as extensively investigated as regions 2.3 and 2.4. The spacial location and role of the conserved regions 3.1 and 3.2 is not clear. However a number of studies have made an attempt to elucidate their role in σ function.

Several studies have indicated that region 3 is involved in core binding [Zhou *et al.*, 1992; Joo *et al.*, 1998] and becomes inaccessible from outside upon the holoenzyme formation [Nagai and Shimamoto, 1997]. Severinov *et al.* [1994] have shown that a site between E508 and M561, which is mostly a *C*-terminal part of region 3, cross-links to the priming ribonucleotide. With σ subunit RNA polymerase is able to localise transcriptional start site (when transcription is initiated from heteroduplex or nicked templates [Fredrick and Helmann, 1997].

Owens *et al.* [1998a] have shown that K496 in region 3.1 is located very close to both strands of the *E. coli lac*UV5 promoter between positions –14 to –31. S517C in region 3.2 mapped close to position –4 of the template strand.

E598K in the primary sigma of *Myxococcus xanthus* (corresponds to E503 of *E. coli* σ^{70}) was reported to be impaired in function because it leads to growth problems and failure to produce an extracellular signal [Davis *et al.*, 1995].

The single amino acid substitutions P504L and S506F in σ^{70} have been shown to compensate for the lack of ppGpp-dependent functions in ppGpp-null strain [Hernandez and Cashel, 1995]. *In vitro* studies have shown that these substitutions also affect the pattern of abortive transcripts differently from those of productive complexes at T7 A1, T5 N25 and T5 $N25_{antiDSR}$ promoters, in a promoter sequence-specific manner [Hernandez *et al.*, 1996]. $E\sigma^{70P504L}$ and $E\sigma^{70P506F}$ form binary complexes with a smaller fraction of open complexes at λP_RAL promoter which suggests that region 3.1 can be involved in melting dsDNA or in maintenance of the open complex [Sen *et al.*, 1998]. The abortive complexes ($E\sigma P_{abort}$) with λP_RAL promoter dissociated faster in the $E\sigma^{70P506F}$ holoenzyme [Sen *et al.*, 1998]. These data suggest that region 3.1 is important for RNAP to choose between either productive or non-productive mode of transcription *in vitro*.

1.1.5. Use of cysteine residues and modification producing free radicals in structural and functional studies of σ^{70}.

Cysteine residues in proteins can be chemically modified by a variety of reagents, which allows studies on structure of proteins, their interaction with adjacent molecules, and conformational changes during the reactions catalysed by these proteins. Such studies have recently been applied to σ^{70}, and they included modification of cysteines with reagent Fe-BABE which is capable of generating hydroxyl radicals upon induction [Owens *et al.*, 1998a; Owens *et al.*, 1998b; Bown *et al.*, 1999; Traviglia *et al.*, 1999], and the

use of fluorescent probes [Callaci *et al.*, 1998; Heyduk and Heyduk, 1999]

σ^{70} contains three cysteine residues at positions 132, 291 and 295, all of which are not crucial for σ function [Narayanan and Krakow, 1983]. They are all located within the non-conserved 245 amino acid domain.

The analysed single-cysteine σ^{70} derivatives demonstrate different reactivity *in vitro*. Owens *et al.* [1998a] and Callaci *et al.* [1998] used $\sigma^{70\Delta cys}$ which is σ^{70} with all three cysteines replaced with serine residues. These substitutions have no negative effects on $\sigma^{70\Delta cys}$ function. Callaci *et al.* reported that *in vitro* activity of $\sigma^{70\Delta cysS438C}$, $\sigma^{70\Delta cysT440C}$ and $\sigma^{70\Delta cysS442C}$ is 37, 59 and 95% of that of $\sigma^{70\Delta cys}$, respectively. Single cysteine substitutions created by Owens *et al.* at different conserved regions of σ^{70}, including S517C in region 3.2, (K376C (2.1), N396C (2.2), R422C (2.3), K496C (3.1), S517C (3.2) retained 70–95% of wild-type transcriptional activity at the *E. coli lac*UV5 promoter [Owens *et al.*, 1998a].

Callaci *et al.* [1998] using σ^{70} with single cysteine replacements analysed cysteine accessibility and fluorescence spectra of fluoroprobe-modified cysteines of free σ^{70} and $E\sigma^{70}$. They showed that upon binding to the core enzyme conformational changes occur at positions 1, 59, 336, 440, 583, 596. Cysteines at positions 440 (T440C, region 2.4), 583 and 596 (T583C and R596C, region 4.2), and 1 become more exposed to solution, whereas T440C and R596C find themselves in more polar environments upon σ^{70} binding to the core enzyme. Cysteines replacing S438 and S442 which belong to the same α-helix as T440 [Malhotra *et al.*, 1996] undergo no significant changes. Cysteines at positions 59 (A59C, region 1.1) and 366 (S366C, nonconserved region) become less exposed, with 59C finding itself in less polar, and 366C in more polar, environments.

1.1.6. The region of σ^{70} between the conserved regions 1.2 and 2.1.

The so called non-conserved region of σ^{70} is a 245 amino acids long stretch of peptide between the conserved regions 1.2 and 2.1. It is absent from all the alternative sigmas, and all of the primary sigmas from gram-positive bacteria analysed so far [reviewed by Lonetto *et al.*, 1992].

The previously analysed single amino acid substitutions in this region do not affect the function of the protein [e.g. Callaci *et al.*, 1998; Callaci and Heyduk, 1998; Owens *et al.*, 1998b]. However the part of the nonconserved region proximal to 2.1, together with part of 2.1, is thought to be essential for core binding [Lesley and Burgess, 1989]. A small deletion in this region (Δaa 330–343, encoded by *rpoD800*, or *rpoD285* allele) results in temperature sensitivity [Liebke *et al.*, 1980; Hu and Gross, 1983]. σ^{70} from *E. coli* C, apart from three substitutions of non-conserved amino acids, differs from σ^{70} from *E. coli* K-12 by a short deletion in the non-conserved region (Δ195–204) [Christie and Cale, 1995].

On the basis of homology of σ^{A} of *Bacillus subtilis* and σ^{70} of *E. coli* Kumar *et al.* [1995b] created $\sigma^{\Delta 245}$ by deletion of amino acid residues 130 through 374 of σ^{70}. The resulting protein reconstituted with the core enzyme exhibits *in vitro* activity indistinguishable from that of σ^{70}, in that it can generate transcripts at 37°C from many different promoters : *trp*, *dnaQ* (P1 and P2), *rnh*, *lac*UV5, *leuX*, *rrnE* (P1, P2 and P_X), *gal* (P1, CRP-dependent, and P2, CRP-independent), although it had significantly reduced ability to transcribe from *alaS* *in vitro* [Kumar *et al.*, 1995b].

Gene *rpoDΔ245* encoding $\sigma^{\Delta 245}$ failed to complement *rpoD285* (ts) *in vivo* at 42°C [Kumar *et al.*, 1995b]. However, *rpoD285* itself encodes σ^{70} with a deletion of aa 330–343 (see above), therefore it was not unexpected that $\sigma^{70\Delta 245}$ would also result in temperature-sensitivity.

In the hybrid sigma test (described in Chapter 2.5.9), the 245 aa deletion was introduced into a hybrid $\sigma^{70/32}$, consisting of 529 *N*-terminal amino acids of σ^{70} followed by 82 *C*-terminal amino acids of σ^{32} [Kumar *et al.*, 1995a]. This hybrid sigma has region 2 from $\sigma^{70\Delta245}$ and region 4 from σ^{32} (heat shock sigma [Grossman *et al.*, 1984]), and was shown to be able to activate a hybrid promoter (σ^{32} consensus for –35, σ^{70} consensus for –10 element) *in vivo*, but less effectively as $\sigma^{70/32}$ hybrid sigma at 30, 37 and 42°C with the activity of $\sigma^{70\Delta245/32}$ shown to be 55, 30 and 24% of that of $\sigma^{70/32}$, respectively.

The nonconserved region is seen as a separate domain in the crystal structure of the σ^{70} fragment, thus confirming the correctly chosen position for 245 amino acid deletion [Malhotra *et al.*, 1996] (Fig. 1.1.4). The *C*-terminus of region 1.2 on helix 1 is in close proximity to the *N*-terminus of region 2.1 on antiparallel helix 12a, and a short amino acid bridge would be enough to connect them without distorting the overall structure of the rest of the fragment.

In protein footprinting experiments, Nagai and Shimamoto [1997] observed that the environment around part of σ^{70} between amino acids 167 and 200, which includes the acidic region, does not change upon σ^{70} binding to the core enzyme. However this part becomes more exposed after the binary complex formation. In the crystal structure of the σ^{70} fragment this stretch of amino acids is partly a loop, with the acidic region being unstructured [Malhotra *et al.*, 1996]. Whether this indicates a significant conformational change in σ^{70} during binary complex formation is unclear.

We decided to study the role of individual amino acids of σ^{70} of *E. coli* in the process of DNA melting by introducing single cysteine residues in several positions of $\sigma^{\Delta245}$, so using it as a source of Cys-less but still functional sigma. If the resulting proteins retain activity *in vitro* and *in vivo*, they will be useful for studying the mechanisms of sigma function. The single cysteines can be specifically modified with either reagents capable of generating free radicals, or with fluoroprobes, which would allow us to study the

proximity of the modified cysteines to different regions of both strands of promoters at different steps of melting, by regulating temperature and salt concentration, or to analyse conformational changes of different parts of the sigma molecule during transcription initiation steps.

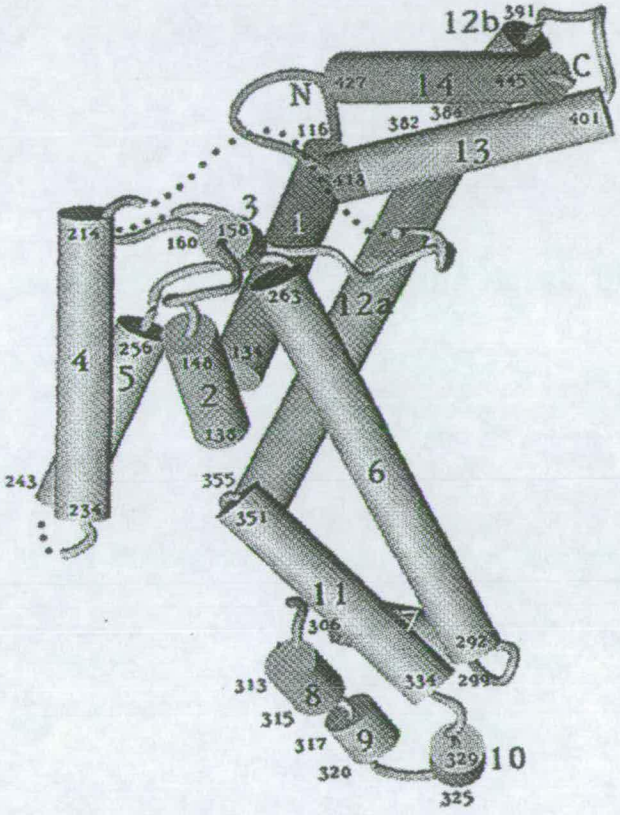


Figure 1.1.4. The structure of a fragment of σ^{70} (residues 114–442) [Malhotra *et al.*, 1996]. The conserved region 1.2 ends in the middle of helix 1, and the conserved region 2.1 begins in the middle of helix 12a.

1.2. *scrP*, an erstwhile candidate minor sigma gene of *Escherichia coli*.

1.2.1. Discovery and early studies of ScrP.

While searching for yet unidentified sigma subunits in *E. coli*, Fujita *et al.* [1987] identified a 27 kDa protein designated ScrP (Sigma Cross-Reacting Protein) as one of several proteins which cross-reacted with antibodies raised against a polypeptide consisting of an amino acid sequence present in the highly conserved region 2.2 of RNA polymerase σ^{70} and σ^{32} subunits. The polyclonal antiserum raised against the synthetic tetradecameric peptide (DLIQEGNIGLMKAV) was tested for cross-reaction with an SDS–PAGE-fractionated whole cell lysate from *E. coli* by Western blotting. It was found that there was cross-reaction not only with known σ^{70} and σ^{32}, but also with several other proteins in the extract, one of which had an apparent molecular size of 27 kDa. Subsequently this was found to be a mixture of two proteins, both of which demonstrated cross-reactivity with the antiserum [Ueshima *et al.*, 1992]. One of these was found to be ribosomal protein S2. The other, ScrP-27A (now abbreviated to ScrP), behaved differently during purification, migrating more slowly on high-resolution SDS–PAGE, and had an *N*-terminal sequence with no similarity to known proteins: MKKIGVILSGXGVYDGSEIHEAVL (X = unknown). More specifically, it was found that immunological cross-reactions between ScrP and anti-σ peptide antibodies were specifically competed out by hexapeptide GLMKAV, which constitutes the carboxyl terminus of the original tetradecameric peptide, while the reaction of σ^{70} with the antiserum was not competed out by GLMKAV. This observation lead the authors to suggest that ScrP should contain a sequence similar, if not identical, to GLMKAV [Ueshima *et al.*, 1992]. However, ScrP, whose sequence is now known, does not contain a amino acid stretch resembling GLMKAV. The nature of the relevant epitope(s) in ScrP remains unknown.

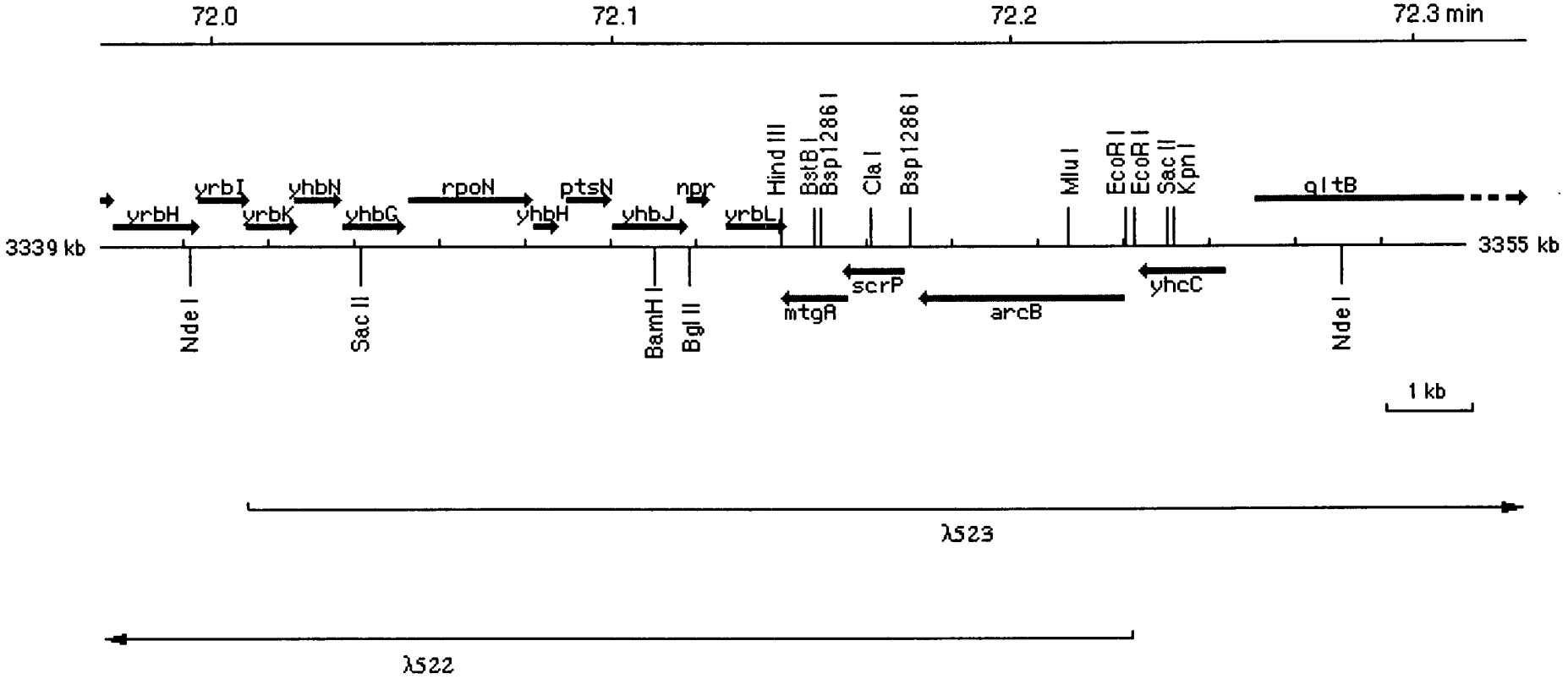


Figure 1.2.1. A part of 72 min of *E. coli* chromosome with the genes, relevant restriction sites and λ phages of Kohara library [Kohara *et al.*, 1987] shown.

After centrifugation of a crude *E. coli* extract (DNase and RNase treated) through a glycerol gradient the 27 kDa proteins (not then distinguished) were found to be distributed throughout all fractions, despite their low molecular size [Fujita *et al.*, 1987]. This was due to either self-aggregation or association with other cellular components. Attempts to detect complex formation between purified ScrP and RNA polymerase core enzyme were unsuccessful, suggesting that this is not a sigma factor [Ueshima *et al.*, 1992].

The map position of the gene, *scrP*, encoding the above protein, was found by reverse genetics. A pair of degenerate oligonucleotides corresponding to the *N*-terminus of ScrP hybridised with a pair of contiguous clones of the Kohara library [Kohara *et al.*, 1987], λ522 and λ523, which overlap by about 9 kb near 72 min on the genetic map, or near 3350 kb on the physical map [R. Hayward, T. Suzuki, N. Fujita and A. Ishihama, unpublished]. The physical location of *scrP* was more precisely defined by restriction and Southern hybridisation analyses [Townsley, 1991]. The closest mapped genes were *arcB* at 3350 kb and *rpoN* at 3342 kb (Fig. 1.2.1.). In order to characterise the *scrP* locus an *Eco*R I – *Hin*d III fragment of λ522 DNA was subcloned into the pJJW11 [Wright and Hayward, 1987]. The resulting recombinant plasmid pFMT1 was thought to carry *N*-terminally truncated *arcB* and full length *scrP.* The minicells with pFMT1 specifically expressed two major polypeptides of about 28 and 14 kDa, as was detected by SDS–PAGE. The 28 kDa protein was assumed to be ScrP; the nature of 14 kDa polypeptide was unclear. It was also determined that the direction of transcription of *scrP* on the chromosome was counter clockwise and therefore the same as for *arcB* [Smillie, 1994].

Sequence analysis of the *scrP* locus identified a good ribosome-binding site Shine-Dalgarno consensus, AGGAuG, situated 6 nt upstream from the initiation codon. Primer extension analysis of total *E. coli* RNA using a primer complementary to 78 through 61 nt of the *scrP* coding sequence generated reverse transcripts terminating at many locations within downstream end of *arcB* and the *arcB-scrP* intergenic region, some of them closest to *scrP*

marked here by arrows. Possible –10 and –35 elements of a σ^{70} dependent promoter were proposed [Smillie, 1994]:

```
GTTACAAGGGCACAGCACTGTTTTTCAGGCCGCACGCAGTTAAGATAGCAAAACTTAA
          ↑                              ↑          ↑

      -35                          -10
ATGATTTGTTACATGAATCAGTTAAATGTGTGATGCGTACCATTTTTATTAAAAAAAT
        ↑                ↑↑       ↑ ↑          ↑↑

TAATGTGTTTCAGTCAATTGCGAAAGGATGATCACAATGAAGAAAATT
                         RBS        ScrP start
```

This preliminary result suggested that *scrP* was cotranscribed with *arcB*, from the *arcB* promoter, but also raised a possibility of independent transcription from a putative promoter(s) within the *scrP–arcB* intergenic region. The other signals by primer extension could reflect mRNA processing sites and/or pause sites in the reverse transcription [Smillie, 1994].

1.2.2. Characterisation of *scrP* by gene disruption.

The genomic copy of *scrP* in two different *E. coli* strains (NM621 and MM38) was mutated by insertion of a kanamycin resistance cassette between 518 and 519 base pair of the *scrP* coding sequence. Cells carrying this disrupted *scrP* did not differ from their isogenic *scrP*$^+$ parents in growth rate either in rich medium (L-broth) or glucose minimal medium. They were able to grow on a variety of carbon sources and at various temperatures between 30 and 46°C, and under anaerobic conditions at 37°C [Smillie, 1994]. The possibility that ScrP is important for Arc function (see below) was also investigated by dye-sensitivity tests [Buxton and Drury, 1983]. The *scrP*-disrupted strains were no more dye-sensitive than their parents [Smillie, 1994].

It had previously been demonstrated that all the genes between *rpoN* and *arcB* (these include *scrP*) were inessential for viability in media supplemented with glutamine, based on studies of strain ECL593 carrying Δ(*zgi*::Tn*10*–*rpoN*) which included *rpoN–arcB* region [Iuchi *et al.*, 1989a].

1.2.3. ScrP protein.

ScrP has a calculated molecular size of 22,982 which is less than the 27 kDa value estimated from its mobility on SDS–PAGE. Hydrophilicity analysis has shown no major hydrophobic or charged regions within the ScrP polypeptide sequence. However, short moderately hydrophobic stretches of amino acid residues near the *N*-terminus might allow ScrP to associate with the cell membrane [Smillie, 1994].

[^{35}S]Methionine labelling of minicells carrying the pFMT1 or its vector parent pJJW11 demonstrated that polypeptides with estimated molecular sizes of about 14 kDa and 28 kDa were encoded only by pFMT1. These polypeptides were presumed to be the products of *scrP*, or restart fragments encoded by supposedly truncated *arcB*, or products of unmapped genes lying downstream of *scrP* and present on pFMT1 [Smillie, 1994]. Preliminary attempts to immunoprecipitate proteins from the minicell extracts with the anti-2.2-peptide antiserum failed to recover the 28 kDa polypeptide, suggesting either a difference in the accessibility of relevant epitopes on ScrP in solution and on blotting membranes, or that the 28 kDa polypeptide produced by pFMT1 was not ScrP, but one of the products of then unmapped genes lying downstream of *arcB* and present on pFMT1.

These early studies failed to cast light on the function of ScrP. Moreover, there was no sequence similarity between ScrP and any other known sequences available in DNA and protein databases, including σ subunits [Smillie, 1994].

Because of the complete lack of similarity between ScrP and the conserved regions of σ subunits, and the apparent absence of interaction between ScrP and RNA polymerase [Ueshima *et al.*, 1992], it was clear that ScrP was not a transcription protein. It was a novel protein with unknown function, not required for *E. coli* survival under normal grown conditions. It was possible to find out more about this protein by finding out if *scrP* expression is co-

regulated with that of the genes neighbouring it on the *E. coli* chromosome, and potentially belonging to the same operon.

1.2.4. Polypeptides with sequence similarities to ScrP.

Recently the sequences of two eukaryotic proteins with a significant similarity to ScrP have been reported. One of these is the ES1 protein from zebrafish, which is specifically expressed in the photoreceptor cells of the retina, but neither in the adult zebrafish brain nor in non-neuronal tissues (including muscle, skin and connecting tissue). ES1 open reading frame encodes a polypeptide of 270 amino acids, with a calculated molecular size of 30.7 kDa [Chang and Gilbert, 1997]. ES1 has 31% identity and 57% overall sequence similarity to ScrP.

There are four types of photoreceptor cells in zebrafish retina: short single cones, long single cones, double cones, and rods. ES1 protein has been detected by immunostaining only in the inner segments of cone, but not in rod, photoreceptor cells (inner segment is outside of outer nuclear layer of the cells) [Chang and Gilbert, 1997]. No staining appeared in the nuclei of photoreceptor cells, and in detergent phase separation experiment ES1 largely partitioned into the aqueous phase, suggesting that it is not an integral membrane protein [Chang and Gilbert, 1997].

ES1 is not detected in zebrafish embryos until they are about 20 days old, when they complete major differentiation processes and begin to become adult-like, although all photoreceptor cells can already be identified at 12 days postfertilisation. Therefore it seems that ES1 protein is only expressed in mature photoreceptors [Chang and Gilbert, 1997].

The 21q22.3 region of human chromosome 21 has been searched for genes responsible for several inherited disorders which have been genetically or physically mapped to this region. As a result three research groups have identified a human gene, variously named KNP-1, GT335, or HES1, mapping to 21q22.3 [Nagamine *et al.*, 1996;

Lafreniere *et al.*, 1996; Scott *et al.*, 1997]. This gene spans an estimated 12–13 kb of genomic DNA and contains seven exons expressed as a 1.7–1.8 kb mature transcript. Some cDNA samples isolated from foetal brain have demonstrated that exon 5 is involved in alternative splicing, resulting in a mRNA species lacking this 93 nt exon and predicting to produce a polypeptide 31 amino acids shorter than the full-length protein [Lafreniere *et al.*, 1996; Nagamine *et al.*, 1996]. The gene is predicted to encode a full-length protein of 268 amino acids, which has 44% identity and 62% similarity to ScrP, and 49% identity and 66% similarity to ES1. The protein itself has not been detected.

The KNP-1/GT335/HES1 mRNA is expressed in various human tissues, but is particularly highly abundant in heart and skeletal muscle [Lafreniere *et al.*, 1996; Nagamine *et al.*, 1996; Scott *et al.*, 1997], and relatively abundant in some tissues involved in the endocrine system [Nagamine *et al.*, 1996].

Potential mitochondrial targeting signals have been found in both the human and the zebrafish proteins, with a possible cleavage site after Val-45 in the human protein. This is consistent with the high expression level in human muscle tissues, the cells of which are rich in mitochondria [Nagamine *et al.*, 1996; Scott *et al.*, 1997].

The similarity of ScrP, ES1 and KNP-1/GT335/HES1 extends throughout their entire sequences, with the exception of the *N*-terminus which is 21 (ES1) or 38 (KNP-1/GT335/HES1) amino acids longer than that of ScrP [Nagamine *et al.*, 1996; Scott *et al.*, 1997; Lafreniere *et al.*, 1996; Chang and Gilbert, 1997]. These differences might be related to different polypeptide targeting systems in eukaryotic tissues.

Since ES1 seems to be more specialised than the human protein, Nagamine *et al.* [1996] have speculated that the zebrafish genome may contain another gene that has a greater similarity to human KNP-1/GT335/HES1 and bacterial *scrP* genes, and the human genome may contain another homologue of the ES1 gene,

functionally more closely related to it. Functions of either of them remains unknown.

1.2.5. Environment of the *scrP* gene on the bacterial chromosome.

Now when the *E. coli* K-12 genome has been fully sequenced [Blattner *et al.*, 1997] we know that four contiguous genes or ORFs, *yhcC*, *arcB*, *scrP* and *mtgA* are orientated in the same direction and located at 72 min on the genetic map of the chromosome. The genes immediately downstream and upstream of these are transcribed in the opposite direction. (Fig. 1.2.1).

The *scrP* gene lies 229 base pairs downstream of *arcB* which encodes a known protein involved in regulation of anaerobic metabolism. Adjacent to *scrP* there is an open reading frame, *mtgA*, whose translational initiation codon overlaps with the translational termination codon of *scrP*, raising a strong possibility of their translational coupling and co-transcription. The translational termination codon of *mtgA* in turn overlaps with the translational termination codon of ORF *yrbL*, which would be transcribed in the opposite direction. If *mtgA* and *yrbL* are transcribed under the same growth conditions, their transcription might interfere with each other, and the complementary downstream ends of their transcripts are likely to interact with each other.

Among other identified genes lying downstream and transcribed in the same direction as *yrbL* are *rpoN*, encoding the alternative sigma subunit σ^{54}, required for transcription of genes controlled by availability of N-sources [Merrick *et al.*, 1993]; *ptsN*, whose product is involved in PTS (phosphotransferase system)-related carbon and organic nitrogen regulation [Powell *et al.*, 1995]; *npr*, which encodes a phosphocarrier protein of the PTS [Powell *et al.*, 1995].

It seems that *arcB* is transcribed from its own contiguous promoter (see later). Thus the putative gene *yhcC* will not be further discussed. However, in order to understand the possible function of

ScrP it is of interest to investigate what is known about the functions of the other two genes which belong to the same putative operon as *scrP*.

1.2.6. The *mtgA* gene and its product.

The *E. coli* genome sequencing project revealed an open reading frame whose beginning overlapped with the end of *scrP* by 4 bp and was orientated in the same direction. The overlap of the stop and start codons and relatively poor agreement of this ORF's ribosome-binding site with Shine-Dalgarno consensus suggest that translational coupling with *scrP* would be likely and perhaps even obligatory. The ORF contains 729 bp and codes for a putative protein of 242 amino acid residues with calculated a molecular size of 27.3 kDa. Comparison of the predicted amino acid sequence with those of other known and putative proteins showed a strong similarity of its product (and an almost identical putative protein from *Klebsiella pneumoniae*) to the *N*-terminal (transglycosylase) region of class A bacterial penicillin-binding proteins represented in *E. coli* by PBP1A and PBP1B. It was therefore suggested that the ORF encodes a monofunctional biosynthetic peptidoglycan transglycosylase, and the name *mtgA* was proposed [Spratt *et al.*, 1996].

Bacterial penicillin-binding proteins (PBPs) are involved in the insertion and cross-linking of new peptidoglycan chains in the murein sacculus. They are membrane-bound proteins and are able to specifically and covalently bind β-lactam antibiotics, such as penicillin, via a mechanism involving an internal serine residue, for example, Ser-465 in the 850 amino acid PBP1A of *E. coli*. The amino-terminal regions of PBPs 1A and 1B form the transglycosylase domains that catalyse glycan chain elongation, while their penicillin-binding *C*-terminal domains are transpeptidases that catalyse peptidoglycan crosslinking. The latter is inhibited by binding of penicillin. The bifunctionality of the PBPs is revealed by the reaction products they generate upon incubation with the lipid-linked precursor GlcNAc-MurNAc pentapeptide

diphosphoryl-undecaprenol [Nakagawa *et al.*, 1984]. *E. coli* mutants lacking either PBP1A or PBP1B are viable, but the double deletion is lethal [e.g. Yousif *et al.*, 1985]. None the other known *E. coli* PBPs have been shown to possess biosynthetic glycosyltransferase activity. On the basis of sequence similarity analysis, recently discovered PBP 1C can be another candidate for bifunctional transglycosylase/transpeptidase enzyme [Schiffer *et al.*, 1997].

Monofunctional transglycosylases (deficient in transpeptidase activities and unable to bind penicillin) have been isolated from *Bacillus megaterium* [Taku *et al.*, 1982], *Staphylococcus aureus*, *Micrococcus luteus* [Park and Matsuhashi, 1984] and *Streptococcus pneumoniae* [Park *et al.*, 1985]. A monofunctional biosynthetic transglycosylase was also isolated from *E. coli* [Hara and Suzuki, 1984]. It had an apparent molecular size of 34 kDa and showed peptidoglycan synthesis activity, i. e. it synthesised chains from disaccharide lipid intermediates UDP-*N*-acetylmuramil-pentapeptide and UDP-*N*-[^{14}C]glucosamine. It did not bind penicillin, and the activity of this enzyme required presence of Ca^{2+} or Mn^{2+}, which could not be replaced by Mg^{2+}. The pH optimum for this peptidoglycan synthesis reaction was between pH 6.0 and 6.5, whereas that of PBP1B, the major transglycosylase in *E. coli*, is between pH 7.5 and 8.0 [Hara and Suzuki, 1984].

Di Berardino *et al.* [1996] cloned and overexpressed *mtgA* from *E. coli* strain W3110. They called the gene *mgt*, for monofunctional glycosyltransferase, to distinguish this glycan-synthesising enzyme from the well known lytic transglycosylases, involved in lysis of peptidoglycan (all five known *E. coli* lytic transglycosylases are listed in Holtje, 1998 #79).

Even when the *mtgA* gene was overexpressed, it was possible to visualise its product on SDS–PAGE only after fractionation of cells into cytoplasmic and membrane fractions. A band corresponding to the expected size of *mtgA* product (~27 kDa) appeared in the membrane fraction. This suggests that the putative transmembrane domain in the *N*-terminal amino acid sequence of MtgA may serve

as a membrane anchor. The W3110 cells stopped growing after induction of *mtgA* overexpression, indicating that the overexpressed gene product is deleterious, although the cells did not lyse for at least 30 minutes [Di Berardino *et al.*, 1996]. It was also demonstrated that the membrane fraction of cells in which *mtgA* was overexpressed showed an up to 7-fold increase in glycan polymerisation activity, with a pH optimum between pH 6.0 and 6.5. This activity was not significantly inhibited by flavomycin, a potent inhibitor of the glycosyltransferase activity of PBPs 1A and 1B [Di Berardino *et al.*, 1996]. It is unclear whether or not *mgt/mtgA* encodes the same monofunctional synthetic transglycosylase as that described by Hara and Suzuki [Hara and Suzuki, 1984]. Both enzymes are not sensitive to flavomycin and have the same pH optimum, but they do not seem to have the same requirements for bivalent cations, and their reported mobility on SDS–PAGE is noticeably different.

Denome *et al.* [1999] attempted to make a comprehensive set of deletion mutants lacking eight murein synthesis proteins (PBPs 1A, 1B, 4, 5, 6 and 7, AmpC, and AmpH) in all possible combinations (apart from double mutants of the genes encoding PBP 1A and 1B), thus creating genetic backgrounds to study the physiological roles of nonessential peptidoglycan-reactive enzymes. In the extreme case where peptidoglycan was still synthesised in the absence of ten PBPs (missing or inactivated PBPs 1A, 2, 3, 4, 5, 6, and 7, AmpC, AmpH, and DacD), the structure of the murein sacculus was still sufficiently strong to prevent lysis in L-broth. The known proteins with biosynthetic transglycosylase function remaining in this case were PBP 1B, PBP 1C and MtgA. These authors also constructed a strain carrying a partially deleted *mtgA* (from *Bst*B I to *Hin*d III), which was viable, as expected from the work of Iuchi *et al.* [1989], but they have not yet reported the construction of multiple mutants lacking any PBPs together with MtgA.

The possible role of MtgA in murein synthesis is unclear. One model for the growth of the murein sacculus suggests that it is synthesised by a hypothetical multienzyme complex, a view based

on observations that many proteins involved in murein synthesis are able to interact with each other *in vitro*. It is suggested that the holoenzyme for cell elongation might consist of a dimer of a transglycosylase/transpeptidase bifunctional enzyme, a dimer of a transpeptidase, a dimer of an endopeptidase, and a lytic transglycosylase. On the other hand, the "constriction" multienzyme complex, necessary for synthesising the septum during bacterial division, would consist of a similar set of enzymes plus a monofunctional transglycosylase [Holtje, 1998]. The actual physiological role of MtgA, and its any possible functional relationship to ScrP, remain unknown, especially as both of them are inessential at least under laboratory growth conditions.

1.2.7. The *arcB* gene and ArcB/ArcA signal transduction system.

Two-component regulatory systems.

ArcB protein is a sensor kinase and ArcA is the response regulator of the Arc (for anoxic redox control, formerly aerobic respiration control) phosphorelay regulatory system. Arc is one of the many the so-called two-component regulatory systems. These signalling systems are widespread in eubacteria, and they allow many different environmental changes to elicit adaptive responses in the bacteria. Homologous pathways have also been identified in other organisms, including archaebacteria [Rudolph and Oesterhelt, 1995; Klenk *et al.*, 1997], yeast *Saccharomyces cerevisiae* [Ota and Varshavsky, 1993; Brown *et al.*, 1994] fungus *Neurospora crassa* [Alex *et al.*, 1996], slime mould *Dictyostelium discoideum* [Schuster *et al.*, 1996; Wang *et al.*, 1996], plant *Arabidopsis thaliana* [Chang, 1993; Kakimoto, 1996; Wilkinson, 1995]. Inspection of the entire nucleotide sequence of *E. coli* reveals that there are at least 30 different sensor–regulator pairs in this organism alone [Mizuno, 1997].

The prototypical two-component pathway is comprised of two proteins — a histidine protein kinase, or sensor kinase, and a

response regulator (Fig. 1.2.2.). The *N*-terminal portion of the sensor kinase usually functions as an input domain, which detects environmental stimuli directly or via upstream receptors. The *C*-terminal portion functions as the transmitter module. This generally consists of approximately 240 amino acids and contains several short blocks of conserved residues. The H1 box includes the histidine residue which serves as the site of autophosphorylation with γ-phosphoryl group of ATP. The G1 and G2 boxes are glycine-rich motifs of nucleotide-binding domains; they are separated by a spacer containing a central F box with as yet unknown function. Box N is involved in autokinase activity [Parkinson and Kofoid, 1992]. The large variable region between boxes H1 and N may contain determinants for specific recognition of cognate receivers [Parkinson and Kofoid, 1992; Stock *et al.*, 1995].

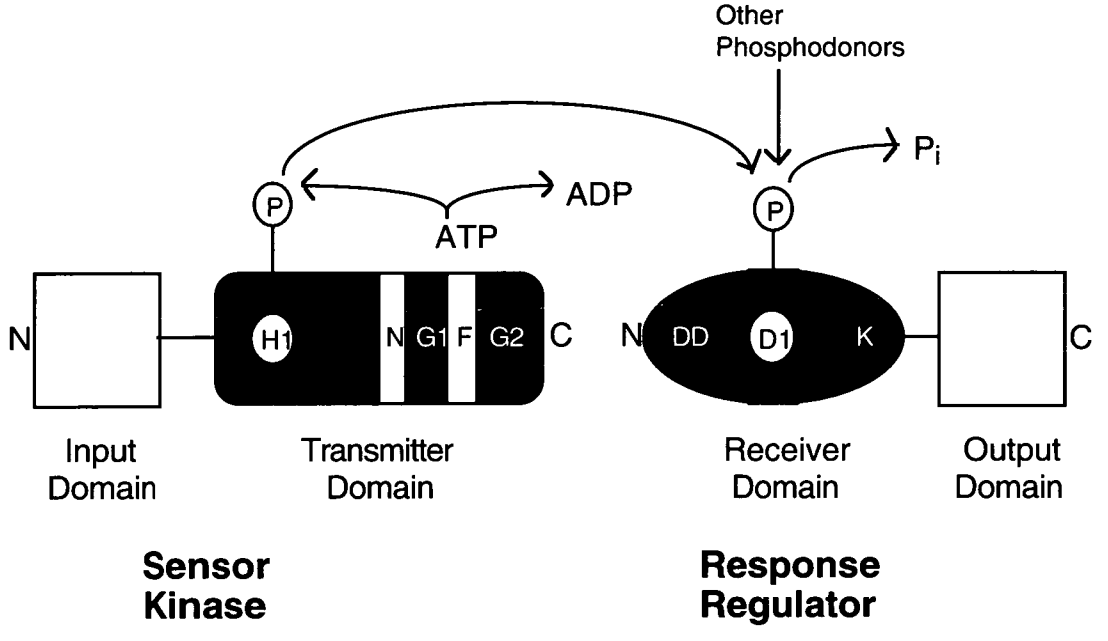


Figure 1.2.2. A prototypical two-component regulatory system with its conserved blocks of residues and transfer of a phosphate group.

The typical response regulator has a receiver domain of about 120 amino acids and contains an invariant aspartate residue (D1) as a phosphorylation site, as well as several other conserved blocks of sequences including an acidic pocket (DD) with the side chain of lysine (K) protruding into it (Fig. 1.2.2.). The structure of the

response regulator CheY is an α/β barrel containing five sets of alternating β strands and α helices connected by short loops or turns, with β strands forming a hydrophobic inner core and α helices wrapped around them, with the residues important for

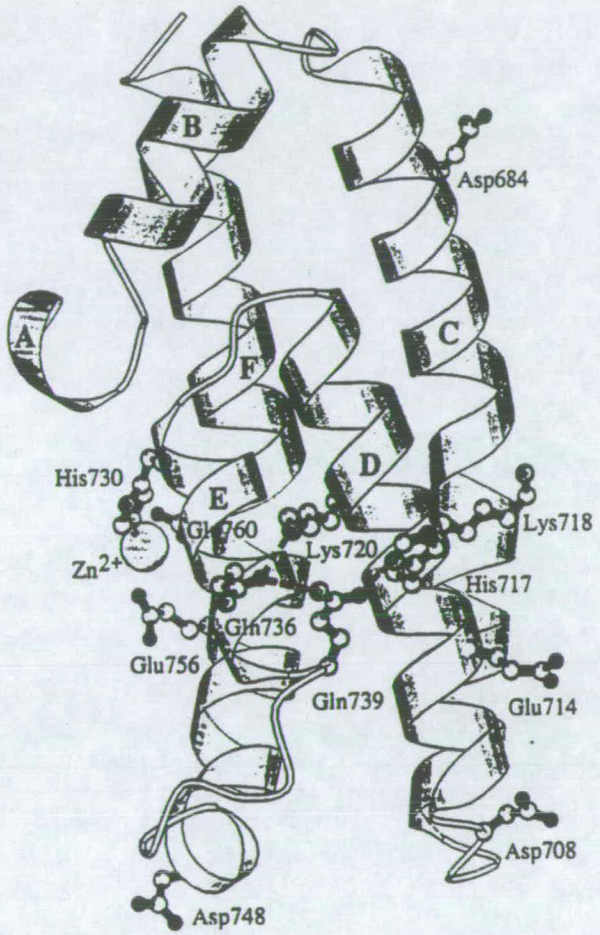


Figure 1.2.3. Structure of the HPt domain of ArcB [Kato *et al.*, 1997], with ribbons representing the main chain folding and some conserved amino acids.

phosphorylation located on the same apical surface [Parkinson and Kofoid, 1992; Volz, 1995]. That this fold may be general for receiver domains has been confirmed by the structures of the nitrate-signalling response regulator NarL from *E. coli*, the sporulation response regulator Spo0F from *Bacillus subtilis*, and the nitrogen response regulator NtrC from *E. coli* [Baikalov *et al.*, 1996; Feher *et al.*, 1997; Volkman *et al.*, 1995]. The receiver domain catalyses transfer of the phosphoryl group from the histidine of the sensor

kinase to the aspartate D1 (Fig. 1.2.2.), and *in vitro* can also utilise a variety of small molecules, but not ATP, as alternative phosphodonors. The phosphorylation state of the receiver domain modulates the activity of the output domain, in bacteria it is usually a *C*-terminal DNA-binding domain which triggers the response to the initial environmental stimulus by appropriate regulation of gene transcription.

Multi-step phosphorelay regulatory systems.

Although the basic sequence of the regulatory pathway is similar for all two-component systems there are many variations, allowing each pathway to suit its specific task. It has been shown that many two-component systems actually comprise four-step phosphorelay (histidine (H1) → aspartate (D1) → histidine (H2) → aspartate (D2)) within two, three or four proteins, some of them containing two or three phosphorelay domains [Appleby *et al.*, 1996] (Fig. 1.2.4.).

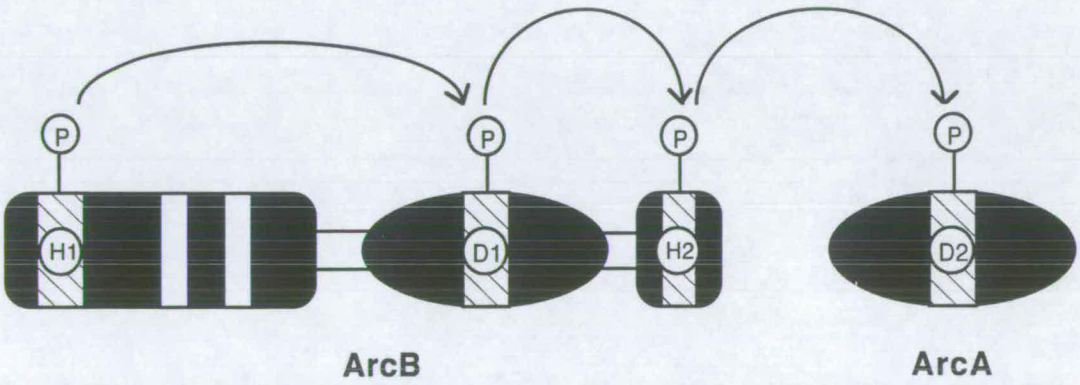


Figure 1.2.4. Four-step phosphorelay within ArcB/ArcA regulatory system. The first three phosphorelay domains are within one protein, ArcB (from [Appleby *et al.*, 1996]).

In one family of multi-step phosphorelay systems the first three steps of phosphorelay occur within a single polypeptide. Examples include ArcB, BarA, EvgS, TorS, and YojN (*E. coli*), BvgS (*Bordetella pertussis*), LemA (*Pseudomonas syringae*), RpfC (*Xanthomonas campestris*), RteA (*Bacteroides thetaiotaomicron*) [Appleby *et al.*, 1996; Parkinson and Kofoid, 1992; Ishige *et al.*, 1994; Matsushika and Mizuno, 1998b]. These proteins contain the first transmitter and receiver domains (H1 and D1) and the second transmitter domain

(H2). The latter, first identified within ArcB, is approximately 120 amino acids long and contains an invariant histidine residue flanked by conserved sequences, but lacks other motifs conserved in the prototypical transmitter domain H1 [Ishige *et al.*, 1994]. This led Mizuno and co-workers to call H2 the unusual histidine phosphotransfer domain, or HPt [Tsuzuki *et al.*, 1995]. X-ray analysis of the isolated HPt part of ArcB revealed a kidney-shaped domain consisting of six α helices and involving an antiparallel four-helix bundle-folding (Fig. 1.2.3) [Kato *et al.*, 1997]. The compact core is predominantly composed of hydrophobic residues contributed to by five of the α helices, with the active His-717 lying in the internal curvature of the kidney-shaped molecule and exposed to the solvent. It has been observed that there is a general shape complementarity between the round surface of the known receiver domains and the concavity of the HPt domain [Kato *et al.*, 1997]. All five of the single amino acid mutations within this domain which were found to impair its function *in vivo* (G685R, E714K, G721D, G724E, as well as H717Y), lie on the same concave surface of the kidney-shaped molecule; moreover, four of these are on the same α helix (Fig. 1.2.3) [Matsushika and Mizuno, 1998b]. Here the replacement of glycines with bulky or charged amino acid residues may hinder the presumed interaction between the HPt domain and the cognate receiver domain.

The ArcB/ArcA signal transduction system.

The ArcB/ArcA system regulates expression of as many as 30 operons in *E. coli* (which comprise the Arc modulon) in response to the redox conditions of growth [Iuchi and Lin, 1988; Lynch and Lin, 1996]. Under anaerobic conditions, ArcA–P (phosphorylated ArcA) is known to repress the expression of at least 17 operons, which contain the genes encoding many aerobic respiratory enzymes, including several dehydrogenases of the flavoprotein class, the cytochrome *o* complex, tricarboxylic acid cycle proteins, enzymes of the glyoxylate shunt and of the pathways for fatty acid degradation. All these enzymes share a common function — they participate in pathways that ultimately donate electrons to molecular oxygen,

with the generation of proton-motive force across the cytoplasmic membrane. ArcA–P also activates at least 9 operons under anaerobic conditions, among them genes encoding enzymes associated with oxygen-limited conditions, such as pyruvate-formate lyase and cytochrome *d*. ArcA–P (but not ArcB – see later) is also required, together with TraJ, to activate the *tra* operon which includes several genes needed to promote conjugative DNA transfer of F-like plasmids [Strohmaier *et al.*, 1998]. ArcA is essential as an accessory protein for Xer site-specific recombination at *psi* to maintain the monomeric form of plasmid pSC101 *in vivo*, but is not absolutely required *in vitro* [Colloms *et al.*, 1998]. With the discovery of the Arc system the term modulon was proposed by Iuchi and Lin [1988] for a group of operons and/or regulons that are under different specific controls, but their expression is modulated by a common pleiotropic *trans*-acting regulatory protein (a global regulator). In contrast to regulons, modulons comprise target operons that do not all share a single specific regulator [Iuchi and Lin, 1993]. However, it is not clear whether to include the *tra* operon in Arc modulon, since ArcB is not involved in its activation (see below) [Silverman *et al.*, 1991a].

The *arcB* gene and its product.

arcB was identified by screening for mutants with elevated anaerobic expression of β-galactosidase by a λ-prophage-borne succinate dehydrogenase transcriptional fusion *sdh–lacZ* in a sdh^+ background [Iuchi *et al.*, 1989a]. The Arc system represses *sdh–lacZ* transcription up to three times even under aerobic, and up to 26 times under anaerobic conditions, judging from its expression in *arcA* and *arcB* [Iuchi and Lin, 1992a; Matsushika and Mizuno, 1998b].

Sequencing of the *arcB* locus revealed an open reading frame of 778 [Iuchi *et al.*, 1990b] or 776 [Blattner *et al.*, 1997], codons (the numbering of amino acids in ArcB in this work is according to Iuchi *et al.* [1990a]). Analysis of the deduced sequence of ArcB suggested that, as discussed previously, the protein belongs to a subgroup of

response sensors consisting of both a transmitter domain and a receiver domain (Fig. 1.2.5). Moreover, it carries two putative transmembrane segments, Phe-23 to Val-50 and Ser-58 to Val-77, to be buried in the plasma membrane with a loop of only 7 amino acid residues in the periplasm. It is not clear whether such a small segment of protein is sufficient to sense a specific external signal. It is possible that either the transmembrane sequences of ArcB are themselves involved in the signal reception or they only anchor the rest of the protein to the cytoplasmic membrane, with stimulus reception then probably occurring near the inner surface of the membrane (in the PAS domain, see below). ArcB has been found in the insoluble cellular fraction, supporting the suggestion that it is a membrane protein. It should be noted that when an *N*-terminally truncated 'ArcB retaining the H1–D1–H2 domains, but lacking the two putative transmembrane regions, was overexpressed in *E. coli*, it was also mainly found in the insoluble fraction. However, this could be an artefact due to the likelihood of overexpressed H1–D1–H2 aggregating in solution [Iuchi *et al.*, 1990b].

Iuchi *et al.* [1990a] showed that ATG(+1) is a real translational start codon for ArcB, and suggested that AGGAauU, 5 nucleotides upstream from it, is the ribosome-binding sequence. They also proposed a putative promoter for *arcB* as shown below. This has a good agreement with –35 and –10 consensus sequences for σ^{70} (marked in bold), and the optimal (17 bp) spacing between:

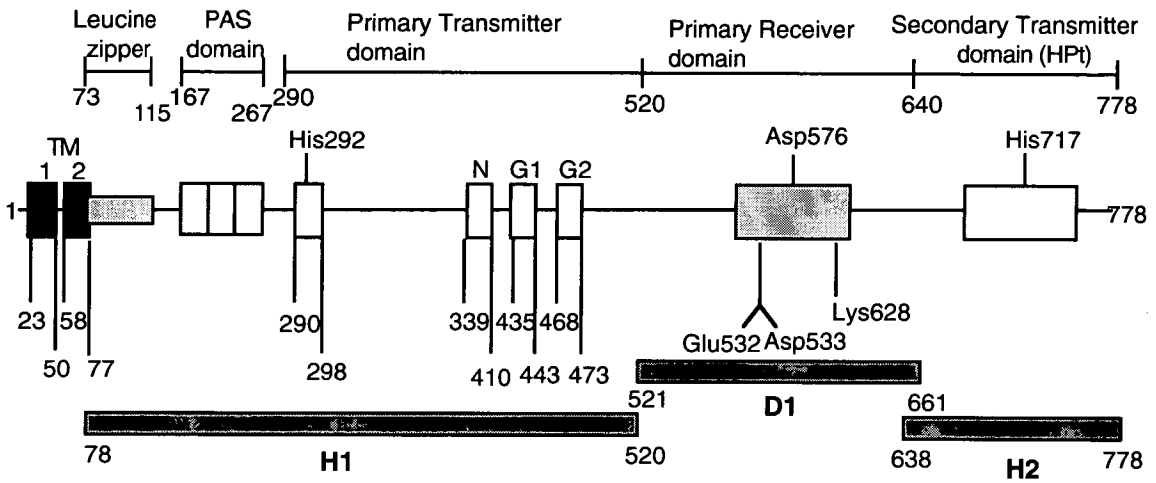
```
end of yhcC              EcoR I
ACG GAG TAGGTCGTTGAGGGGAATTCCGCATTTCTCACACAATTTATAACGTA
Thr Glu *                                          -35

                                                EcoR I
ACTGTCAGAATTGGGTATTATTGGGGCAGGTTGTCGTGAAGGAATTCCCTAATG...
                -10                          RBS       +1
```

The *arcB* promoter is likely to be in the *Eco*R I–*Eco*R I region, because the *Kpn* I–*Bsp*1286 I DNA fragment containing a part of *yhcC* and the full length *arcB* cloned on a multicopy plasmid under a weak plasmid promoter resulted in slow growth of host cells, presumably due to overproduction of ArcB, which is a membrane

protein. Deletion of the *Eco*R I–*Eco*R I region resulted in normal growth rate and still good yield of ArcB [Iuchi and Lin, 1992a]. This is either due to the destroying of *arcB* promoter, resulting in transcription of *arcB* only from the weak plasmid promoter; or due to creating a weaker ribosome-binding site for *arcB*, gGGAauU, instead of AGGAauU.

ArcB sensor kinase protein



ArcA response regulator protein

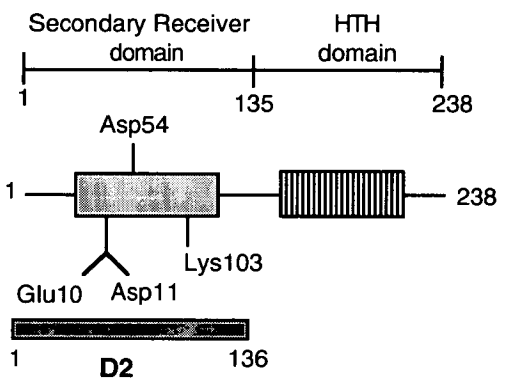


Figure 1.2.5. Schematic representations of ArcB and ArcA. The two transmembrane segments (TM) were predicted on the basis of hydrophobicity plot [Iuchi *et al.*, 1990b]. The sensor PAS domain was identified by Taylor and Zhulin [1999]. Some conserved amino acids in transmitter and receiver domains and catalytic determinants of the first transmitter domain are indicated. H1, H2, D1 and D2 show polypeptides used in studies *in vitro* described further.

They also suggested that the GC-rich sequence which immediately follows the open reading frame of *arcB* might be a Rho-independent terminator:

```
..ACT.AAA AAA TGACCCCGGCTAGACCGGGGTGCGCGAATAC..
  Thr Lys Lys *
  776 777 778
```

This proposal is not very plausible, because there is no U-rich RNA sequence characteristic of such terminators downstream of the stem-loop [e.g. d'Aubenton-Carafa, 1990]; moreover, unless *arcB* is poorly translated, ribosomes terminating the polypeptide might inhibit stem-loop formation [e.g. Wright, 1987]. If the stem-loop does from, its function might rather be to protect the *arcB* mRNA from 3'-exonuclease attack [McLaren *et al.*, 1991]. In fact the sequence is a very good candidate for Rho-independent termination of transcription converging on *arcB* from downstream, because the DNA sequence encoding the lysine codons would produce a U-rich tail downstream of the stemloop in any complementary RNA. This terminator might be required to protect significant upstream progression of RNA polymerase into *arcB*, originating from an unknown promoter between *arcB* and *scrP*, or even from the genes distal to *mtgA* (Fig. 1.2.1).

To study if the expression of *arcB* changes itself during anaerobiosis, Iuchi and Lin [1992b] constructed a strain carrying a λ-prophage-borne *arcB–lacZ* fusion. The *lacZ* expression appeared to be a little higher under aerobic (mid-exponential phase) than under anaerobic (non-shaken overnight cultures) growth conditions, with the levels of β-galactosidase activity remaining essentially the same whether or not $ArcB^+$ function was provided by a complementing plasmid.

Possible signals for Arc system activation.

Available data indicate that signals to ArcB may come both from the membrane and the cytoplasm. D-Lactate, pyruvate, NADH and acetate, which are cellular metabolites whose levels are

increasingly accumulated during anaerobiosis, stimulate ArcB phosphorylation levels in membrane preparations *in vitro* [Iuchi, 1993]. It has also been shown that the generation of the signal activating Arc system *in vivo* requires terminal cytochromes, suggesting that the stimulus sensed by ArcB could be a terminal electron carrier in a reduced form that accumulates in the absence of O_2 [Iuchi *et al.*, 1990a]. Possible candidates include the lipid-soluble electron carriers ubiquinone and menaquinone [Lynch, 1996]. Another suggestion is that a source of signalling for ArcB is the proton potential across the membrane, which is high during aerobic respiration. ArcB would thus act as a protonmeter [Bogachev *et al.*, 1993; Lynch, 1996]. Taylor and Zhulin [1999] identified a PAS signal module within the cytoplasmic part of ArcB (Fig. 1.2.5.). The PAS sensor domains are found in a variety of proteins involved in signal transduction, are located inside cells and sense redox changes in the electron transport system or overall cellular redox status. Therefore it seems likely that PAS domain of ArcB senses a cytoplasmic signal(s). Activation of PAS domain should bring conformational changes to H1 domain of ArcB; another possibility is that it causes dimerisation of PAS domains of two neighbouring ArcB proteins, as it was observed with some other proteins possessing PAS domains [Taylor and Zhulin, 1999].

Aerobic repression of Φ(*sdh–lacZ*) by ArcB has been observed in cells grown in L-broth or minimal media supplemented with glucose, galactose or xylose. However, in minimal media with acetate, succinate or fumarate (intermediate products of the aerobic respiration pathway) as the source of carbon and energy, the aerobic repression of Φ(*sdh–lacZ*) is completely eliminated, such that the levels of its expression in an *arcB*$^+$ and in an *arcB* null mutant become indistinguishable. It has also been observed that the derepressed levels of Φ(*sdh–lacZ*) expression in *arcB*$^-$ strain differs for different sources of carbon and energy. Thus factors other than the ArcB/ArcA system must also be implicated in the regulation of *sdh* under the observed conditions [Matsushika and Mizuno, 1998a], and it is not clear whether acetate, succinate or fumarate can be signals for ArcB.

The *arcA* gene and its product.

Mutations in *arcA*, also referred to as *cpxC* (for conjugative plasmid expression), *dye* (for sensitivity to redox dyes), *fexA* (for F expression), *msp* (for male-specific phage), *seg* (for segregation of F-plasmid), or *sfrA* (for sex factor regulation), were initially identified as having impairment of sex pilus synthesis in F^+ or Hfr cells, which also made them resistant to male-specific bacteriophages which use the pilus for attachment [Buxton and Drury, 1983; Buxton and Drury, 1984]. Iuchi and Lin [1988] then showed that in *arcA* mutants the levels of various enzymes associated with anaerobic metabolism were not increased, under anaerobic conditions, in the normal way. At the same time phosphorylated ArcA acts as activator for another set of genes, such as *cydAB* (encoding cytochrome *d* oxidase) and *pfl* (pyruvate formate lyase) under anaerobic and microaerobic conditions [Lynch, 1996]. It was later shown that the anaerobic regulation and sex factor regulation functions of the *arcA* gene product are separable. A strain with the *sfrA5* mutation (producing ArcA with a V203M substitution in its *C*-terminal region) is impaired in expression of *tra* genes, but shows normal repression of *sdh* during anaerobiosis, whereas *arcB* mutants and the *arcA1* mutant (with an 8 amino acid insertion in the *N*-terminal region of ArcA) show derepressed *sdh* expression, but have only slightly reduced levels of pilus production and *tra* operon expression [Silverman *et al.*, 1991a; Iuchi *et al.*, 1989b; Iuchi and Lin, 1992a]. There is also evidence that anaerobiosis only slightly affects *traY* promoter activity [Silverman *et al.*, 1991b].

It was proposed that the second regulator of ArcA function, which is implied by the above evidence, could be the *cpxA* gene product, which could serve as sensor for the sexual state of the cell [Iuchi *et al.*, 1989b]. However, it has recently been established that another gene, *cpxR*, encodes the cognate response regulator for the CpxA sensor protein, and this two-component system is involved in protein trafficking in *E. coli* envelope [Pogliano *et al.*, 1997].

By utilising *arcA–lacZ* transcriptional and translational fusions it has been found that both transcription and translation of the *arcA* gene during anaerobic growth are about four times higher than that during aerobic growth. Both ArcA and Fnr (for fumarate and nitrate reduction), another anaerobic global regulator, are required for full expression of Φ(*arcA–lacZ*), Fnr has been found to bind and activate one of at least two *arcA* promoters [Compan and Touati, 1994].

Phosphorelay within the Arc system.

The conserved amino acids taking part in the ArcB/ArcA phosphorelay are His-292 (H1), Asp-576 (D1) and His-717 (H2) of ArcB and Asp-52 (D2) of ArcA (Fig. 1.2.5.). The importance of His-292 and Asp-576 in ArcB has been demonstrated *in vivo*, when the plasmid-borne *arcB* (H292Q), *arcB* (H292L), *arcB* (D576Q) mutant alleles failed to repress either aerobic or anaerobic expression of Φ(*sdh–lacZ*) in strains carrying *arcB* null or *arcB1* (a 1.3 kb insertion which introduces a stop-codon at *arcB* codon 241) on their chromosomes. *arcB* (D533A) and *arcB* (D576A) mutant alleles also failed to repress Φ(*sdh–lacZ*) anaerobically. Plasmid-borne alleles containing only transmitter domain H1, as well as ArcB^{H717L}, have repressed aerobic expression of the fusion to a degree similar to that of the wild type [Iuchi and Lin, 1992a; Matsushika and Mizuno, 1998a].

The ArcB^{H292Q} or ArcB^{H292L} proteins show essentially no autophosphorylation *in vitro* [Iuchi, 1993; Tsuzuki *et al.*, 1995], while ArcB^{D533A}, ArcB^{D576A} and ArcB truncated at amino acid 516 (ArcB^{tr516}) show reduced autophosphorylation [Iuchi, 1993], whereas ArcB^{D576Q}, somewhat surprisingly, exhibits very good autophosphorylation *in vitro* [Tsuzuki *et al.*, 1995].

The importance of His-717 was first demonstrated by Ishige *et al.* [Ishige *et al.*, 1994] when they showed that the ArcB and BarA proteins, as well as their HPt domains alone, can act as multicopy suppressors of the *ΔenvZ* mutation, whereas HPt domain carrying the H717L replacement cannot. However, although ArcB^{H717L} does

not repress L-lactate dehydrogenase (one of the enzymes under Arc control) anaerobically, it is nevertheless dye-resistant. This led to the suggestion that H2 might be only an alternative to H1 phosphodonor to H1 [Ishige *et al.*, 1994]. Later studies *in vitro*, however, showed that H2 is in fact the major phosphodonor for ArcA (see ArcA Phosphorylation *In Vitro*).

Autophosphorylation of ArcB .

The H1–D1–H2 polypeptide (*N*-truncated ArcB) rapidly undergoes autophosphorylation, in which ^{32}P-labelling occurs from [γ-^{32}P]ATP but not from [α-^{32}P]ATP or [γ-^{32}P]GTP [Iuchi and Lin, 1992b]. H1–D1–H2 (and also to some extent ArcB^{tr516}) can then transphosphorylate ArcB^{H292Q}, but not the double mutant ArcB$^{H292Q,\ D576A}$ [Iuchi, 1993]. Georgellis *et al.* [1997], using purified individual ArcB/ArcA transmitter and receiver domains (H1–D1–H2, H1–D1, D1–H2, H1, D1, H2 and ArcA) in various combinations, showed that only those polypeptides containing H1 (with His-292) undergo autophosphorylation. The kinetic analysis has demonstrated that initial rates of autophosphorylation of the H1–D1 and H1–D1–H2 polypeptides are very rapid, reaching a peak and subsequently declining, whereas labelling of the isolated H1 is slow and increases continually throughout the experiment. This suggests that the D1 domain assists in releasing phosphate from the H1 and/or the D1 domains.

The purified HPt (H2) domain of ArcB is not capable of autophosphorylation, but can be phosphorylated intermolecularly by ArcB–P, ArcB^{D576Q}–P, ArcB^{H717L}–P, although not by ArcB^{H292L}–P [Ishige *et al.*, 1994; Tsuzuki *et al.*, 1995].

The study with purified ArcB/ArcA domains showed that the D1 domain is significantly more active than ArcA as an acceptor of phosphoryl group from the H1–P domain and that ArcA in turn is significantly more active than H2 [Georgellis *et al.*, 1997]. It was also shown that the phosphorylated H1–D2 polypeptide rapidly transphosphorylates H2.

Phosphorylation of ArcA *in vitro*.

ArcA is phosphorylated by ArcB–P, but does not undergo autophosphorylation in the presence of ATP [Iuchi and Lin, 1992b; Georgellis *et al.*, 1997]. Tsuzuki *et al.* [1995] were the first to demonstrate direct phosphotransfer from the isolated H2 domain to D2 *in vitro*, i. e. when ArcA rapidly acquired phosphate from the phospho-HPt domain. Iuchi [1993b] showed that truncated ArcB–P possessing only the transmitter domain H1–P ($ArcB^{tr516}$–P) can phosphorylate ArcA at a low level, whereas ArcB–P with mutated aspartates in the receiver domain ($ArcB^{D533A}$–P and $ArcB^{D576A}$–P) do not transphosphorylate ArcA at all. However, Tsuzuki *et al.* [Tsuzuki *et al.*, 1995] have shown that ArcA can be phosphorylated by $ArcB^{D576Q}$–P almost as efficiently as by ArcB–P, and much less efficiently by $ArcB^{H717L}$–P. These data are consistent with the results of phosphorylation of ArcA by a mixture of the H1–P and H2–P domains: firstly ArcA rapidly receives a phosphoryl group from H2–P, and then slowly from H1–P [Georgellis *et al.*, 1997]. In another experiment a mixture of H1–P and D1 slowly phosphorylated ArcA, but after addition of H2 to the reaction ArcA phosphorylation rapidly increased [Georgellis *et al.*, 1997]. They have also shown that phospho-H1–D1 barely phosphorylates ArcB, while phospho-H1–D1–H2 and phospho-H1–D1 with the addition of H2 are capable of doing so.

The apparently contradictory behaviour of $ArcB^{D576A}$ and $ArcB^{D576Q}$ is presumably due to the nature of the amino acid replacements. Asp-576 is assumed to be the crucial active residue, since it is conserved amongst receiver domains in other phosphorelay systems. This assumption is confirmed by the behavior of the $ArcB^{D576A}$ mutant protein. Perhaps in $ArcB^{D576Q}$ the acidic pocket, including among other conserved residues Asp-533, is to some extent still able to accept the phosphoryl group. Another possible explanation is that phosphotransfer in $ArcB^{D576Q}$ is carried out by an H1→H2 route, bypassing the D2 domain. This reaction was observed to some extent *in vitro* by Georgellis *et al.* [1997], when fragments of normal ArcB were used.

Based on the kinetics of autophosphorylation of H1–D1–H2, H1–D1 and H1 described above, Georgellis *et al.* [1997] and Tsuzuki *et al* [1995] have proposed models for signal transduction in the Arc system which is summarised in the figure 1.2.6 *a*.

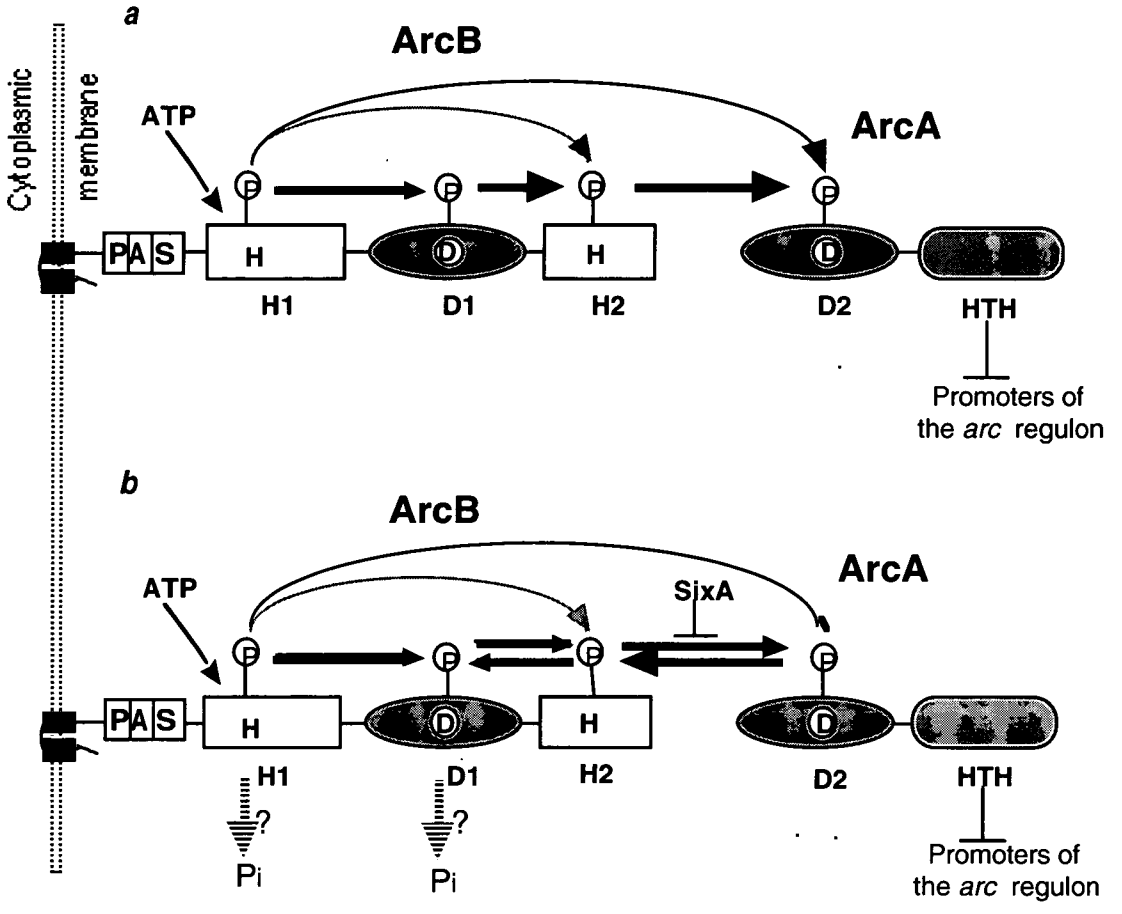


Fig. 1.2.6. *a*) Model for signal transduction by the Arc system. Arrows show the directions of phosphoryl group transfer along alternative pathways, with the thickness of arrows indicating the relative activity of these pathways observed in studies *in vitro*. *b*) Signal transduction and signal decay in the ArcB/ArcA phosphorelay system. The reverse phosphorelay pathway is based on the data of Tsuzuki *et al.* [Tsuzuki *et al.*, 1995], Georgellis *et al.* [Georgellis *et al.*, 1998], and Ogino *et al.* [Ogino *et al.*, 1998].

Signal decay in the ArcB/ArcA phosphorelay system.

The phosphate from H2–P can be transferred *in vitro* to another site on ArcB and ArcB^{H717L} rapidly and transiently, and to a lesser extent to ArcB^{H292L}, although the latter reaction is not transient.

ArcB^{D576Q} also exhibits some transient ability to acquire phosphate from the phospho-HPt domain [Ishige *et al.*, 1994; Tsuzuki *et al.*, 1995].

ArcB domains H1–D1–H2, D1–H2 or a mixture of D1 and H2 rapidly catalyse dephosphorylation of ArcA–P and of ArcA'–P (the isolated D2 domain of ArcA), and H1–D1–H2 and D1–H2 undergo only transient phosphorylation during these reactions. Tested separately, H1, H1–D1, D1 or H2 polypeptides do not exhibit ArcA'–P dephosphorylating activity and do not acquire detectable levels of phosphoryl group from ArcA'–P [Georgellis *et al.*, 1998]. Thus it has been concluded that D1 and H2 each provide an essential catalytic function for the ArcA'–P dephosphorylation process (Fig. 1.2.6.).

D1–H2^{H717Q}, a mixture of D1^{D576A} and H2 or a mixture of D1 and H2^{H717Q} do not exhibit significant phosphatase activity towards ArcA'–P. However, when ArcA'–P is incubated with D1^{D576A}–H2, the phosphoryl group is transferred to the mutant polypeptide. It was therefore concluded that ArcA'–P dephosphorylation proceeds by a reversal direction of the normal His→Asp→His→Asp phosphorelay, except that H1 seems to be dispensable [Georgellis *et al.*, 1998]. On the other hand, the observations that phosphotransfer from H2–P to ArcB^{H292L} is non-transient [Tsuzuki *et al.*, 1995], suggest that the H1 domain is expected to be involved in release of the phosphate from D1–P. Further work need to be done to confirm which residue – H1 or D1 – releases phosphate from ArcB/ArcA in the reverse phosphorelay (Fig. 1.2.6.).
Effectors, possible candidates for an input signal for ArcB activation, found to accelerate the rate of autophosphorylation of ArcB (D-lactate, acetate, pyruvate), as well as ATP, show no significant effect on dephosphorylating activity of H1–D1–H2 and D1–H2 on ArcA'–P [Georgellis *et al.*, 1998]. The object of the possible action of effectors may be H1 domain, provided it is not involved in phosphatase activity of ArcB. They also may affect only the direct, but not reverse, phosphorelay.

SixA as a specific phosphatase for ArcB.

As mentioned previously, *arcB* (or its HPt-encoding fragment) can act as a multicopy suppressor of an *ΔenvZ* deletion. Moreover, ArcB–P can phosphorylate OmpR, the cognate regulator of EnvZ, *in vivo* (see next chapter). Recently Ogino *et al.* [1998] identified, cloned and expressed *sixA* (for signal inhibitory factor-X), a gene whose presence on a plasmid eliminates multicopy *arcB* suppression of *ΔenvZ*, presumably by preventing His–Asp phosphorelay between the HPt domain of ArcB and the receiver domain of OmpR. SixA contains an arginine–histidine–glycine motif, including a crucial histidine, which is found in a number of phosphatases. Purified SixA can rapidly dephosphorylate the phospho-HPt domain of ArcB. During this reaction no phosphorylated SixA is detected. SixA exhibits no such *in vitro* activity with the phosphorylated transmitter domain (H1) of EnvZ. These data allowed the authors to suggest that SixA can function as a phosphohistidine phosphatase that may affect the H2–D2 phosphorelay through regulating the phosphorylation state of the HPt domain of ArcB [Ogino *et al.*, 1998]. SixA has been crystallised recently, and its structure should soon be resolved [Hamada *et al.*, 1999].

Based on all the known data about ArcB/ArcA signal transduction the scheme described in Fig. 1.2.6. can be proposed.

Does "cross-talk" between members of non-cognate phosphorelay families occur under physiological conditions?

It has been reported that receivers can acquire phosphoryl groups not only from their cognate transmitters, but also from heterologous transmitters *in vivo* and *in vitro*, and directly from small molecules such as, in case of ArcA, acetylphosphate or carbamoylphophate *in vitro* [Drapal and Sawers, 1995; Lynch and Lin, 1996]. However, most *in vitro* and *in vivo* observations of such "cross-talk" have involved protein stoichiometries not relevant under physiological conditions [Parkinson and Kofoid, 1992].

In case of ArcB, there have been reports of such cross-talk between ArcB and OmpR or CheY, the cognate response regulators of EnvZ and CheA, respectively. As mentioned previously, the genes *arcB* and *barA* are multicopy suppressors of the *ΔenvZ* mutation. The corresponding phosphoproteins *in vitro* can phosphorylate OmpR, which is the cognate regulator of EnvZ [Ishige *et al.*, 1994]. CheY and OmpR have been shown to compete for interaction with ArcB, when both ArcB (or its HPt domain) and CheY are overexpressed in Δ*envZ* cells carrying Φ(*ompC–lacZ*). Moreover the phosphorylated HPt domain of ArcB can efficiently transfer its phosphoryl group to CheY *in vivo* and *in vitro* [Yaku *et al.*, 1997].

However, in single copy ArcB and ArcB^{H717L} are indistinguishable in respect of OmpC and OmpF production in *envZ*$^{+}$ background, which suggests that under physiological conditions ArcB does not play any role in OmpR regulation [Ishige *et al.*, 1994; Tsuzuki *et al.*, 1995]. Similarly, Iuchi and Lin [1992b] showed that ArcB and ArcB^{tr516} supplied on a multicopy plasmid do not significantly affect expression of Φ(*ompF–lacZ*), Φ(*ompC–lacZ*), or Φ(*narC–lacZ*). This suggests that ArcB does not communicate with OmpR or NarL, the response regulators of other two-component regulatory systems, when their cognate sensors (EnvZ and NarX, respectively) are functioning normally.

Finally, Georgellis *et al.* [1997] compared the ability of the two ArcB phosphoryl transmitters to donate phosphoryl groups to ArcA and to the non-cognate response regulators PhoB and CpxR. H1–D1–P was incubated with ArcA, PhoB and CpxR for a short time in the presence or absence of H2. Significant phosphorylation occurred only with ArcA and only when H2 was present. Therefore, neither H1 nor H2 of ArcB efficiently recognises its non-cognate response regulators PhoB and CpxR.

Thus, considering the available data on ArcB and other transmitter-containing proteins, the cross-talk between non-cognate transmitters and receivers is probably of little physiological significance.

CHAPTER 2

MATERIALS AND METHODS

2.1. Bacterial strains, phages, plasmids, oligonucleotides and proteins.

Bacterial strains used in this study are listed in Table 2.1. Bacteria were stored in frozen storage buffer at –70°C.

Bacteriophages used in this study were P1 and M13mp18–14 (M13mp18 with its *Sac* I–*Pst* I segment replaced with the *Sac* I–*rpoD–HgiA* I fragment of pSAK3 [Kumar *et al.*, 1993]). P1 phage lysates were stored at 4°C as broth suspensions to which a few drops of chloroform had been added to prevent microbial growth. M13 phage infected bacteria (strain TG1) were stored in frozen storage buffer at –70°C.

Plasmids used and constructed in the course of this study are listed in Table 2.2.

Table 2.1. Bacterial strains.

Strain	Genotype	Source/Reference
AF681	MC4100 *rpoS*::Kan^R *λ(uspA–lacZ)*	[Farewell *et al.*, 1996]
BL21(λDE3)	*E. coli* B F$^-$ *dcm ompT hsdS* (r_B^- m_B^+) *gal lon* λ(DE3)	Laboratory (Mishima) stock [Studier and Moffatt, 1986]
BL21(λDE3) pLysS	*E. coli* B F$^-$ *dcm ompT hsdS* (r_B^- m_B^+) *gal lon* λ(DE3) [pLysS Cm^R]	Laboratory stock [Studier *et al.*, 1990]
BW313	*thi-1 relA1 spoT1? dut-1 ung-1* HfrKL16(PO45)	Laboratory stock [Kunkel, 1985]
DA26	*ΔarcB*::Cm^R *araB1 Δ(lac–pro)*	S. McAteer [Ishige *et al.*, 1994]
DH5α	F$^-$ *supE44 lac*ΔU169 *hsdR17 recA1 endA1* *ϕ80(lacZ*ΔM15*) gyrA96* (Nal^R) *thi-1 relA1*	Laboratory stock [Hanahan, 1983]

Strain	Genotype	Source/Reference
DS941	AB1157 *recF* $lacI^{Q}\Delta$M15 (F^{-} *thr-1 leu-6 thi-1 kalK2 ara-14 xyl-5 proA2 his-4 argE3 supE44* Str^{R})	S. Colloms [Colloms *et al.*, 1998]
DS941 *arcB*::Tn*5* 5.4		S. Colloms [Colloms *et al.*, 1998]
DS941 *arcA*::Tn*5* 2.3		S. Colloms [Colloms *et al.*, 1998]
ECL590	MC4100 sdh^{+} *(sdh–lacZ*$^{+}$*) arcB1 Δfrd-1*	S. Colloms [Iuchi *et al.*, 1989]
ECL1601	F^{-} *araD139 Δ(argF–lac)U169 rpsL150 relA1 flb5301 deoC1 ptsF25 rbsR Δ(frd)101* sdh^{+} *sdh–lacZ*$^{+}$	S. Colloms [Iuchi and Lin, 1988]
MC4100	F^{-} *araD139 Δ(argF–lacZ)U169 rpsL150 relA1 flb5301 deoC1 ptsF25 rbsR*	[Casadaban, 1976]
MM20*asn*$^{+}$	F^{-} *argG6 hisG1 leuB6 metB1 pyrE gal-6 lacY1 xyl-7 supE44 uhp* bgl^{+} *fhuA2* ($T1^{R}$) ($\phi 80^{R}$) *gyrA* (Nal^{R}) *rpsL104* (Str^{R}) *tsx-1* ($T6^{R}$) *polA* (Amp) λ^{-}	S. McAteer
MM38	F^{-} *argG6 hisG1 leuB6 metB1 pyrE gal-6 lacY1 xyl-7 supE44 uhp* bgl^{+} *fhuA2* ($T1^{R}$) ($\phi 80^{R}$) *gyrA* (Nal^{R}) *rpsl104* (Str^{R}) *tsx-1 arcB* λ^{-}	S. McAteer
NM621	F^{-} *hsdR* (r_K^{-} m_K^{+}) *mcrA mcrB supE44 recD1009*	Laboratory stock [Whittaker *et al.*, 1988]
P678.54	*thr leu thi minB*	S. McAteer
TA6	TP8503 *Δ('arcB–scrP–mtgA–yrbL–'npr)::(arcB`–'lacZ, CAT).* TP8503 transduced with P1(TA5) (37°C). Cm^{R}, Amp^{S}.	This work (Fig. 4.12)
TA8	TP8503 *arcA2 zjj*::Tn*10 arcB–lacZ CAT*	This work
TA9	TP8503 *arcA*::Tn*5* 2.3 *arcB'–'lacZ CAT*	This work

Strain	Genotype	Source/Reference
TA10	TP8503 *zjj::*Tn*10 arcB–lacZ CAT*	This work
TG1	F' [*traD*36 *proAB*$^+$ *lacI*^{Q} *lacZ*ΔM15] *thi-1 supE hsdΔ5 Δ(lac-proAB)*	[Sambrook *et al.*, 1989]
TM5	MM20 transduced with P1(TP8503 [pMLC3]), Cm^R (30°C).	This work
TM6	TP8503 *Δ('mtgA–yrbL–'npr)::(mtgA`–'lacZ, CAT).* TP8503 transduced with P1(TM5) (37°C). Cm^R, Amp^S.	This work (Fig. 4.12)
TP8503	*thi-1 leu thr supE42 Δ(lac–pro) fhuA* Tn*7*, (tmp^R)	S. McAteer
TP8503 *arcA2 zjj::*Tn*10*		This work
TS5	MM20 transduced with P1(TP8503 [pSLC3]), Cm^R (30°C).	This work
TS6	TP8503 *Δ('scrP–mtgA–yrbL–'npr)::(scrP`–'lacZ, CAT).* TP8503 transduced with P1(TS5) (37°C). Cm^R, Amp^S.	This work (Fig. 4.12)
TS6 *rpoS*	TS6 *rpoS*::Kan^R Cm^R	This work
TS9	MM20*asn*$^+$ transduced with P1(TP8503 [pSLK3]), Kan^R Amp^S Cm^S (30°C).	This work (Fig. 4.12)
TS10	TP8503 *Δ('scrP–mtgA–yrbL–'npr)::(scrP`–'lacZ,* Kan^R) *ΔarcB::*CAT	This work (Fig. 4.12)
TS11	TP8503 *Δ('scrP–mtgA–yrbL–'npr)::(scrP`–'lacZ,* Cm^R) *arcB::*Tn*5* 5.4	This work (Fig. 4.12)
TS12	TS6 *zjj::Tn10 arcA2* Cm^R Tet^R	This work
TS13	TS6 *zjj::*Tn*10* Cm^R Tet^R	This work

Strain	Genotype	Source/Reference
TS14	TP8503 *arcA*::Tn*5* 2.3 Δ('*scrP–mtgA–yrbL–'npr*)::(*scrP'–'lacZ CAT*) Cm^R Kan^R	This work

Table 2.2. Plasmids.

Plasmid	Description	Source/reference
pACE1	5.34 kb *Bam*H I–*Eco*R I fragment from the λ523 clone of the Kohara library inserted into the pJJW11.	A. Elding (Fig. 4.2) [Elding, 1995]
pALC1	*Asc* I–*Bgl* II fragment of PCRed *lacZ*–CAT cassette cloned into pBRU12 digested with *Mlu* I and *Bgl* II.	This work (Fig 4.11)
pALC3	*Aoc* I–*Mfe* I portion of pSLC3 cloned into pALC1.	This work (Fig 4.11)
pBR322	Amp^R, Tet^R, pMB1 replicon.	[Bolivar *et al.*, 1977]
pBRU12	*Eco*R I–*Bam*H I fragment of pACE1 cloned into pUC18.	This work (Fig. 4.11)
pDSA28	Insertion of Kan^R (from Tn*903*) cassette into the *Nsi* I site of *scrP* in pFMT1 (antiparallel to *scrP*).	D. Smillie (Fig. 4.2) [Smillie, 1994]
pFMT1	*Eco*R I – *Hin*d III fragment of λ522 clone of the Kohara library (*arcB–scrP–mtgA*) inserted into the pJJW11.	Laboratory stock (Fig. 4.2) [Townsley, 1991]
pHW111	A hybrid $\sigma^{70/32}$ donor, under control of P_{lac}. *lacI*^{Q} (divergent from P_{lac}), Rep_{p15A}, Kan^R (Tn*5*).	Laboratory stock [R. Hayward and H. S. Williamson, unpublished]

Plasmid	Description	Source/reference
pJJW11	Derivative of pHR9. *Eco*R I – P_{gal} – *Hind* III –*Bam*H I followed by (promoterless) *galK*. Amp^R.	Laboratory stock (Fig. 4.2) [Wright and Hayward, 1987]
pML_{HM}::*lacZ*	Hybrid promoter plasmid with regions –35 for σ^{32} and –10 for σ^{70}. *lacZ* reporter. Amp^R, pMB1 replicon.	Laboratory stock [A. Kumar]
pMLC1	*Bst*B I–*Bgl* II fragment of PCRed *lacZ*–CAT cassette cloned into pBRU12 digested with *Bst*B I and *Bgl* II.	This work (Fig 4.11)
pMLC2	pALC1 with *lacZ'* sequence (of pUC18) removed.	This work (Fig. 4.11)
pMLC3	*Aoc* I–*Mfe* I portion of pSLC3 cloned into pMLC1.	This work (Fig. 4.11)
pSAK15–70/32	A hybrid $\sigma^{70/32}$ donor, under control of P_{lac}. Rep_{p15A}, Amp^R.	Laboratory stock [Kumar *et al.*, 1995a]
pSLC1	*Bst*B I–*Bgl* II fragment of PCRed *lacZ*–CAT cassette cloned into pBRU12 digested with *Cla* I and *Bgl* II.	This work (Fig. 4.11)
pSLC2	pALC1 with *lacZ'* sequence (of pUC18) removed.	This work (Fig. 4.11)
pSLC3	*Aoc* I–*Nde* I portion of *lacZ* in pSLC2 replaced with that from pML_{HM}::*lacZ*.	This work (Fig. 4.11)
pSLK3	pSLC3 with *Hind* III–*CAT*–*Bam*H I replaced with *Hind* III–Kan^R–*Bam*H I from pUC4K.	This work (Fig. 4.11)
pTAK1	*Hind* III–*Bam*H I fragment from pACE1 replaced with a synthetic *Hind* III – *Not* I – *Bgl* II linker (primers TK1 and TK2).	This work (Fig. 4.2)

Plasmid	Description	Source/reference
pTAK2	Insertion of Kan^R cassette (*Acc I* fragment from pUC4K) into the *Bst*B I site of *mtgA* in pTAK1 (parallel to *mtgA*).	This work (Fig 4.4)
pTAK3	The same as pTAK2, but Kan^R is antiparallel to *mtgA*.	This work (Fig 4.4)
pTAK4	Insertion of Kan^R cassette (*Acc I* fragment from pUC4K) between the *Bst*B I sites of *mtgA* and *galK* in pTAK1 (antiparallel to *mtgA*).	This work (Fig 4.4)
pTAK5	The same as pTAK4, but Kan^R is antiparallel to *mtgA*.	This work (Fig 4.4)
pTAK6	Insertion of Kan^R cassette (*Acc I* fragment from pUC4K) into the *Bst*B I site of *galK* in pTAK1 (parallel to *galK*).	This work (Fig 4.4)
pTAK7	Overproducer of fusion protein P7-'ScrP and MtgA. *Xmn* I–*Hin*d III fragment of pTAK1 was cloned into pT7-7 digested with *Sma* I and *Hin*d III. Amp^R.	This work (Fig 4.4)
pTAK8	Overproducer of fusion protein P8-'ScrP and MtgA. *Nsi* I–*Hin*d III fragment of pTAK1 cloned into pT7-7 digested with *Pst* I and *Hin*d III. Amp^R.	This work (Fig 4.6)
pTAK9	Overproducer of ScrP and MtgA. *Mfe* I–*Hin*d III fragment of pTAK1 cloned into the pT7-7 vector digested with *Eco*R I and *Hin*d III. Amp^R.	This work (Fig 4.6)
pTAK11	Overproducer of ScrP and MtgA. *Bfa* I–*Hin*d III fragment of pTAK1 cloned into the pT7-7 vector digested with *Nde* I and *Hin*d III. Amp^R.	This work (Fig 4.6)

Plasmid	Description	Source/reference
pTK111	*Bam*H I–*Xho* I regions of pHW111 replaced with a *Bam*H I–*Not* I–*Xho* I linker (oligonucleotides N4356 and N4357).	This work (Fig 4.1)
pTK113–pTK424	Described in Table 4.1.	This work
pTK210	*Xba* I–*Not* I–*Nsi* I fragment of pGEMDΔ245 (in front of *rpoD*$^{\Delta 245}$) is replaced with *Xba* I–*Sac* I–*Nsi* I (oligonucleotides RH7 and RH8).	This work (Fig 4.1)
pTK211	*Bam*H I–*Xho* I regions of pTK210 replaced with a *Bam*H I–*Not* I–*Xho* I linker (oligonucleotides N4356 and N4357).	This work (Fig 4.1)
pTK310	*Xba* I–*Not* I–*Nsi* I fragment of pGEMD (in front of *rpoD*) is replaced with *Xba* I–*Sac* I–*Nsi* I (oligonucleotides RH7 and RH8).	This work (Fig 4.1)
pTK311	*Bam*H I–*Xho* I regions of pTK310 replaced with a *Bam*H I–*Not* I–*Xho* I linker (oligonucleotides N4356 and N4357).	This work (Fig 4.1)
pUC4K	ColE1 derived replicon (pMB1). A source of Kan^R (Tn*903*) cassette. Amp^R.	[Vieira and Messing, 1982]
pUC18	ColE1 derived replicon. P_{lac} promoter region, polylinker and *lacZ'*. 2.7 kb. Amp^R.	[Vieira and Messing, 1982]

Oligonucleotides.

5'P_i denotes 5' phosphorylated oligonucleotide.

083V 5'-CTGATGCATCGCTTGTGC-3'

287M 5'-GGCACAATCAACGCATC-3'

876L 5'-CAACGTCAACACCGCTTC-3'

877L 5'-GAAGCGGTGTTGACGTTG-3'

834T 5'-ATGTGTGATGCGTACCAT-3'

997M 5'-CTGCGCCAACACCAGGGA-3'

3'Cml 5'-CGAAGCTT*AGATCT*CAGGCGTTTAAGGGCACCAATAACTGCCT-3'
Bgl II

839F 5'-ACTGACTGAAATGCC-3'

Asc I
ASCI-LACZ 5'-AC*GGCGCGC* CTG GCC GTC GTT TTA CAA CGT C-3'
7 8 9 10 11... (codons of LacZ)

BstB I
Bscrp 5'-GTC*TTCGAA* CTG GCC GTC GTT TTA CAA CGT CG-3'
7 8 9 10 11... (codons of LacZ)

BstB I
Bmtga 5'-CTG*TTCGA ATG* ATT ACG GAT TCA CTG GCC GTC G-3'
3 4 5 6 7 ... (codons of LacZ)

C308 5'-GCAGAACTTTTGCTTCACGC-3'

CATout2 5'-CGAAGTGATCTTCCGTCACAGGTA-3'

ExtP 5'-CTTTGACGTGCGCTTCTTGCC-3'

G0752 5'-CACCGCGGACGAAGATGCTGC-3'

G0754 5'-GTTCCTTGACCTGATTCAGG-3'

IntP 5'-TGCGATGAGTGGCAGGGCGG-3'

N4356 5'-GATCCGTCAGGCGGCCGCATAATAGC-3'

N4357 5'-TCGAGCTATTATGCGGCCGCCTGACG-3'

NS10 5'P_i-GCGTAGGTGCAGAACTTGTAACC-3'

NS11 5'-CGGATCCACGCGGTTGCGTAGGTGCAGAACTTGTAAC-3'

NS12 5'P_i-GGAGAACTTGCAACCACGGC-3'

NS13 5'P_i-CGGATCCACGCGGTTGCGTAGGTGGAGAACTTGCAACCACGGC-3'

NS14 5'P_i-CACGGCGGCATTCGAATTTATCAACC-3'

RH1 5'-CCTGGT*GGATCC*GT*TGC*GCGATCACCCGC-3'
BamH I 437

RH2 5'-GGT*GGATCC*GTCAGGCGATC*TGC*CGCTCTATCG-3'
BamH I 440

RH3 5'-GGT*GGATCC*GTCAGGCGATCACCCGC*TGT*ATCG-3'
BamH I 442

RH4 5'P_i-GGAGATTGGCTCGCAGGCGATCTTCAGC-3'

RH5 5'P_i-CCCCAGATGGCAATCTTCATCATCACC-3'

RH6 5'-GTCGGTCAGACGGTTTGCCGCC-3'

RH7 5'-CTAGTCGTATGAGCTCCGCACCTGCA-3'

RH8 5'-GGTGCGGAGCTCATACGA-3'

TK1 5'-GATCTCAGCGGCCGCTCACA-3'

TK2 5'-AGCTTGTGAGCGGCCGCTGA-3'

Univ.P 5'-TCCCAGTCACGACGT-3'

Proteins.

Restriction endonucleases were purchased from Boehringer Mannheim, New England Biolabs and Promega. DNA modification enzymes used were Sequenase® Version 2.0 T7 DNA Polymerase were (USB), Vent DNA Polymerase (New England Biolabs), T4 DNA Ligase (Boehringer Mannheim and New England Biolabs), Taq DNA Polymerase (Promega). RNase and Proteinase K were from Sigma. Core RNA polymerase and σ^{70} were purified and kindly provided by N. Shimamoto. $\sigma^{70\text{-HMK}}$ and $\sigma^{\Delta 245\text{-HMK}}$ were purified and kindly provided by H. Nagai. Other σ subunits were purified according to the protocol below.

2.2. Growth media and buffers.

L-broth and L-agar were used routinely for most bacterial manipulations. For phage P1 2.5 mM $CaCl_2$ was added to the medium. In certain cases M9 minimal medium, supplemented with appropriate carbon sources, vitamins and amino acids, was used.

2.2.1. Growth media.

L-broth	Difco bacto tryptone	10 g
	Difco bacto yeast extract	5 g
	NaCl	5 g
	pH to 7.2 with NaOH	
	Distilled water	to 1 litre
L- agar	L-broth + 15 g Difco agar per litre	
LB top agar	L-broth + 6.5 g Difco agar per litre	
M9 minimal medium	4 x M9 salts	100 ml
	20% carbon source	4 ml
	Supplements as required	
	Distilled water	to 400 ml

M9 minimal agar	As M9 minimal medium + 15 g Difco agar per litre	
4 x M9 salts	NaH_2PO_4	28 g
	KH_2PO_4	12 g
	NaCl	2 g
	NH_4Cl	4 g
	Distilled water	to 1 litre
4 x M9 salts	$MgSO_4.7H_2O$	4 g
	Citric acid	40 g
	KH_2PO_4	400 g
	$NaNH_4.HPO_4.4H_2O$	70 g
	Distilled water	to 1 litre

2.2.2. Commonly used buffers.

Phage buffer	Na_2HPO_4	7 g
	KH_2PO_4	3 g
	NaCl	5 g
	$MgSO_4$ (0.1 M)	10 ml
	$CaCl_2$ (0.1 M)	10 ml
	1% gelatine solution	1 ml
	Distilled water	to 1 litre
Bacterial buffer	KH_2PO_4	3 g
	Na_2HPO_4	7 g
	NaCl	4 g
	$MgSO_4.7H_2O$	2 g
	Distilled water	to 1 litre

TAE buffer

Working solution:

40 mM Tris–acetate
2 mM EDTA

50x Conc. stock solution:

Tris base	242 g
Glacial acetic acid	57.1 ml
0.5 M EDTA (pH 8.0)	100 ml
Distilled water	to 1 litre

TBE buffer

89 mM Tris–borate
89mM boric acid

5x Conc. stock solution:

Tris base	54 g
Boric acid	27.5 g
0.05 M EDTA (pH 8.0)	20 ml
Distilled water	to 1 litre

TE buffer

10 mM Tris–HCl (pH 8.0)
1 mM EDTA (pH 8.0)

2.2.3. Minimal medium supplements.

Amino acid supplements were stored in stock solutions of pure amino acids at a concentration of between 2–10 mg ml^{-1} depending upon the solubility of the particular amino acid. Sparingly soluble amino acids, such as tyrosine, were dissolved in 0.01 M NaOH. The final concentration of the amino acids in the media was usually in the order of 20–100 µg ml^{-1}. If a rich minimal medium was required, vitamin-free casamino acids (CAA) were used. The stock concentration of CAA was 100 mg ml^{-1} and the final concentration in the medium was typically 1–5 mg ml^{-1}. It should be noted that casamino acids lack tryptophan and this should therefore be added to CAA medium if the bacterial strain to be used is auxotrophic for this amino acid.

Purines and pyrimidines were added to minimal media when required. Thymine and uracil were stored at a concentration of 2 mg ml^{-1} in water, and their final concentration in minimal medium was usually 20–40 μg ml^{-1}.

The only vitamin supplement found necessary in the entire course of this work was thiamine hydrochloride (vitamin B1). This was stored as a 1 mg ml^{-1} solution in water and its final concentration in minimal medium was 2 mg ml^{-1}.

2.2.4. Growth of bacteria.

Bacteria were routinely grown as liquid cultures at 37°C (or 30°C for temperature-sensitive strains). Usually fresh overnight cultures that had been inoculated from a single colony were diluted back the following day and grown as required.

To screen colonies for β-galactosidase activity, cells were plated on L-agar containing X-Gal (5-bromo-4-chloro3-indolyl-β-galactoside) at 40 μg ml^{-1}.

For β-galactosidase activity assays in the *scrP* studies L-broth was boiled, cooled down, glucose (to 40 mM) and antibiotics (where necessary) were added. Glucose (as an electron acceptor) is necessary for growth of bacteria under anaerobic conditions, and it was used both in aerobic and anaerobic experiments. Into this media bacteria were diluted 10000-fold from fresh overnight cultures. Aerobic cultures were then grown as usual with vigorous shaking, while for anaerobic growth sealable bottles were filled to top with the liquid cultures, and nitrogen was bubbled continuously through a needle in the seal. A syringe was used to take 1 ml samples through the needle.

2.2.5. Selection of antibiotic resistance.

The concentrations for the antibiotics used in this work are shown in Table 2.3. All antibiotics were used in both complex and minimal media at the same concentrations.

Table 2.3. Antibiotic solutions.

Antibiotic	Abbreviation	Solvent	Concentration	
			Stock (mg ml^{-1})	Final (µg ml^{-1})
Ampicillin	Amp	50% ethanol	100	100
Chloramphenicol	Cm	Ethanol	35	35
D-cycloserine		H_2O	15	150
Kanamycin sulphate	Kan	H_2O	50	50
Tetracycline hydrochloride	Tet	50% ethanol	10	10
Rifampicin	Rif	Ethanol	50	50

2.2.6. Dye sensitivity tests [Iuchi *et al.*, 1990].

Fresh colonies were restreaked on dye agar and incubated overnight at 37°C. After this dye-sensitive fells formed no or tiny colonies, whereas dye-resistant cells formed colonies of usual size.

Dye agar

1% tryptone
0.8% NaCl
0.2 mg ml^{-1} toluidine blue *o* or methylene blue
15 g Difco agar per litre

2.3. Bacterial and phage techniques.

2.3.1. Preparation of competent cells and transformation with plasmid DNA using DMSO.

To prepare competent *E. coli* cells the method of Chung *et al.* [1989] was employed. Five ml of L-broth was inoculated with a single colony and incubated overnight with shaking at a suitable temperature. This culture was diluted 1 in 50 into fresh L-broth and grown with good aeration to an A^{600} of between 0.3 and 0.4. The culture was chilled on ice, transferred to a universal bottle and centrifuged at 3600 x g for 10–15 min. The supernatant was removed and the bacterial pellet resuspended in 0.1x the original volume of ice-cold TSS buffer. At this point the cells could be frozen at –70°C, or could be used immediately for transformation. Freshly prepared cells always gave the highest transformation efficiency. The plasmid DNA (typically 1–100 ng in <10 µl) was added to 0.1 ml aliquots of the competent cells, mixed gently and stored on ice for 15–30 min. After this time 1 ml of LBG (L-broth + 20 mM glucose) was added and the cells incubated with aeration at an appropriate temperature for 1 h to allow expression of plasmid antibiotic-resistance genes. Two hundred microlitres of this mixture was then spread onto antibiotic-containing plates and incubated until bacterial colonies appeared. Whenever a transformation was performed an aliquot of competent cells lacking plasmid DNA was used as a control.

Cells can be transformed with replicative form M13 DNA using this method. In this case the LBG step is omitted and 0.25 ml of M13-sensitive plating cells added to the transformed cells. The mixture is added to 3 ml of molten L-top agar, mixed gently and poured onto a L-agar plate which, once set, is incubated at 37°C overnight, after which M13 plaques should be clearly visible.

TSS buffer:

Difco Bacto Tryptone	10 g
Difco Yeast Extract	5 g
NaCl	10 g
PEG 3350	100 g
$MgSO_4$	20 mM
DMSO	50 ml
PIPES Buffer pH 6.5	10 mM
distilled water	to 1 litre

2.3.2. Preparation of competent cells and transformation with plasmid DNA using calcium chloride.

1 litre of L broth was inoculated with 1 ml of a fresh overnight culture and incubated with good aeration until the A^{600} reached 0.4. The culture was chilled on ice for 15 min. The cells were pelleted in at 4000 rpm in a Sorwall GS3 rotor at 4°C for 5 min. The pellets were resuspended in 1/10 of the initial volume of ice-cold 0.1 M $CaCl_2$ and stored on ice.

To 100 µl aliquots 1–10 ng of plasmid DNA or 5–20 µl of ligation reaction was added, and the mixture was incubated on ice for 1–1.5 hours. After heat shock for 1.5–2 min in a water bath at 42°C the transformation mixture was put on ice for 5 min, then 1 ml of LBG was added to it, and the tube was incubated for 1 hour at 37°C with shaking to allow expression of antibiotic genes. Aliquots of 100–500 ml were plated on L-agar with appropriate antibiotic selection and the plates were incubated overnight at 37°C (or 30°C where stated).

2.3.3. Frozen storage of bacterial strains.

Strains of *E. coli* can be conveniently stored at –70°C without suffering a dramatic loss of viability, this includes strains harbouring plasmids that might otherwise be lost. A fresh 5 ml overnight culture was prepared with antibiotic selection if required. This was centrifuged at 4000 rpm for 10–15 min, the supernatant discarded and the cells resuspended in 0.1x the original volume of

Frozen Storage Buffer. The cells were then left on ice for a couple of hours before storing at –70°C.

Frozen storage buffer:

50% bacterial buffer
50% glycerol (v/v)

2.3.4. Preparation of phage P1 plate lysates.

Cells were grown in L-broth with 2.5 mM $CaCl_2$ until late log phase. 1 ml of the culture was then mixed with 10^6 phage, incubated at 37°C for 30 min, and 3 ml of L-top agar containing 2.5 mM $CaCl_2$ was added. This was poured onto a L-agar plate, left to set and incubated at 37°C overnight or until visible lysis occurred. The layer of top agar was scraped off into a sterile universal bottle. Equal volume of L-broth with 2.5 mM $CaCl_2$ and 100 µl of chloroform were added to the bottle. The bottle was shaken gently for 15 min at 30°C, centrifuged 5 min at 5,000 rpm at room temperature. The supernatant was transferred to a fresh bottle and a few drops of chloroform were added. The lysate was stored at 4°C.

2.3.5. Phage P1-mediated transduction.

The recipient strain was grown up to late-log phase in L-broth. The cells were harvested by centrifugation and the bacterial pellet resuspended in 0.1x the original volume of L-broth with 2.5 mM $CaCl_2$. 100 µl aliquots of the concentrated culture were mixed with either 100 µl or 100 µl of a 10 x dilution of the phage stock. These were incubated at 37°C for 30 min. 1 ml of phage buffer was added, the cells were centrifuged and resuspended in 0.6 ml of LBG. This was incubated at 37°C for 1 h to allow expression of the antibiotic resistance and then 0.2 ml aliquots were plated out on L-broth agar plates containing the appropriate antibiotic. Plates were incubated at a suitable temperature until colonies appeared.

2.3.6. Phage P1-mediated transfer of plasmid DNA to chromosome.

To transfer DNA cloned in pUC-based plasmid the phage P1-mediated transfer protocol was used [Draper *et al.*, 1998]. The DNA to be transferred had a selection marker (Cm^R or Kan^R) attached to it, and was flanked on each side by at least 1 kb of DNA homologous to chromosome. A *rec*$^+$ strain (TP8503) was transformed with a pUC-based plasmid with both (Amp and Cm, or Amp and Kan) antibiotics selection. A P1 lysate was prepared on a purified transformant and this was used to transduce MM20 or MM20*asn*$^+$ strain to Cm^R or Kan^R. MM20 is a *polA* strain and plasmid DNA can not replicate in it. Transductants were selected only in presence of the second antibiotic (Cm or Kan), purified, and one of Amp^S colonies was used to prepare a second P1 lysate. This lysate was in turn used to transduce TP8503 strain to Cm^R or Kan^R. Chromosomal structure of the resulting purified transductants was checked by PCR analysis and by Southern hybridisation.

2.3.7. Purification of minicells for *in vivo* protein synthesis studies [Stoker *et al.*, 1983].

Three 500 ml batches of L-broth containing appropriate antibiotics were inoculated with a culture of minicell strain P678.54, transformed with the required plasmid, and were grown overnight with vigorous shaking. At the same time eight sucrose gradients were prepared by freezing 35 ml aliquots of 20% (w/v) sucrose in M9 glucose medium in 50 ml polycarbonate centrifuge tubes (–70°C) and allowing the frozen solutions to thaw undisturbed at 4°C overnight.

Parental cells were removed from the culture by centrifugation at 2500 rpm in a Sorvall GS-3 rotor for 15 min at 4°C. Minicells were harvested from the supernatant by centrifugation at 8500 rpm in a GS-3 rotor for 20 min at 4°C. The minicell pellets were resuspended in a total of 10 ml of supernatant and the suspensions were transferred to a 30 ml Corex tube on ice. The suspension was

whirlimixed vigourously 3 times for 1 minute intervals to disperse aggregated minicells. 2.5 of minicell suspension was layered onto each of 4 x 35 ml sucrose gradients and these were centrifuged at 5000 rpm in a Sorvall HB-4 rotor for 20 min at 4°C. The minicell bands were collected from the upper part of the gradients and pooled samples from pairs of gradients were pelleted by centrifugation in a Sorvall SS-34 rotor at 13000 rpm for 10 min at 4°C. The resulting minicell pellets resuspended in a pooled total of 4 ml M9 glucose medium, and 2 ml were layered onto each of 2 glucose gradients. The minicell bands were harvested from the gradients as before and the minicells were pelleted. The pellet was resuspended in 2 ml M9 glucose, layered onto a sucrose gradient and centrifuged. The minicells were pelleted in a Sorvall SS-34 rotor at 8000 rpm for 10 min and resuspended in 1 ml of 30% (w:v) glycerol in M9 glucose medium. The A^{600} of a 0.1 ml aliquot diluted to 1 ml was measured and the suspension was aliquoted into Eppendorf tubes so that the volume diluted to 1 ml gave A^{600} 0.2, which is equivalent to approximately 2×10^9 minicells. The cells were then stored at –70°C.

2.3.8. Labelling of plasmid-encoded proteins in minicells.

The minicell preparations were thawed at room temperature and diluted to 1 ml in M9 glucose. The cells were then spun in a microfuge for 2 minutes at 4°C. The pellet was resuspended in 0.1 ml M9 glucose and the samples were incubated for 1 hour at 37°C to remove any bacterial DNA. 20 µCi [^{35}S]methionine was added (this was prepared by diluting [^{35}S]methionine five-fold in 25% (w/v) filter-sterilised Difco methionine assay mix in M9 glucose), and samples were incubated at 37°C for 3 hours. Samples were chased with 5 µl cold methionine (8 mg ml^{-1}) for 3 minutes and then spun for 2 minutes in a microcentrifuge. The pellets were washed with 1 ml 0.05 M Tris–HCl pH 6.8, then resuspended in the residual liquid. 15 µl loading buffer containing β-mercaptoethanol was added, and the samples were fractionated by SDS–PAGE. Labelled proteins were detected by autoradiography.

2.3.9. Labelling of plasmid-encoded proteins in maxicells [Close and Rodriguez, 1982].

Cultures of plasmid-containing CSR603 cells were grown in 3 ml of M9, CAA medium at 37°C with shaking until A^{600} reached 0.4. A 2.5 ml portion of this culture was transferred to a sterile 100-mm diametre petri dish and irradiated for 15 seconds with 254 nm ultraviolet lamp at an intensity of 1.25 J m^{-2} s^{-1}. 2 ml of irradiated cells was transferred into a tube, D-cycloserine was added to a final concentration 150 µg ml^{-1}, and the cells were incubated for 16 hours at 37°C with shaking. Cells from 1.5 ml of his culture were then pelleted, washed twice with M9 medium and resuspended in 0.8 ml of M9 medium. The culture was incubated at 37°C without shaking, after which 0.2 µl M9 medium containing 5 µCi [^{35}S]methionine (1380 Ci $mmol^{-1}$, Amersham) was added. After another hour of incubation at 37°C the cells were pelleted, washed twice with 1 ml of 100 mM NaCl, resuspended in SDS–PAGE loading buffer and separated on 12% SDS–PAGE.

2.4. DNA Techniques.

Most DNA manipulations were performed according to Sambrook *et al.*, [1989] with some modifications, or according to instructions of manufacturers of certain reagents/kits.

2.4.1. Plasmid preparation.

All preparations of plasmid DNA were performed using a modification of the alkaline lysis method of Birnboim and Doly [1979]. When possible, DNA was prepared from an *endA* strain, typically DH5α, to avoid DNA degradation by the EndA endonuclease.

Large scale plasmid preparations.

A single colony of the plasmid-carrying bacterial strain was inoculated into 5 ml of L-broth with the appropriate selection and

incubated overnight at appropriate temperature (typically 37°C) with vigorous shaking. One millilitre of this culture was then used to inoculate 25–500 ml of L-broth, with similar selection, in a 2 litre flask, which was then incubated at 37°C overnight, again with vigorous agitation. The culture was chilled on ice for 15 min, transferred to two 250 ml centrifuge bottles and cells were collected by centrifugation at 6,000 x g in a Sorvall GSA or GS-3 rotor for 10 min at 4°C. The bacterial pellets were then each washed in 100 ml of TE buffer, pooled to give a total volume of 200 ml, and recentrifuged as above. The DNA usually was isolated by using QIAGEN® Plasmid Midi Kit. The resultant cell pellet was resuspended in 4 ml (or in a bigger volume, according to the manufacturer instructions) of cold QIAGEN buffer P1 containing 50 mM Tris–HCl, pH 8.0; 10 mM EDTA; 100 μg ml^{1} RNase A and transferred into a 50 ml centrifuge tube. An equal volume of QIAGEN buffer P2 (200 mM NaOH, 1% SDS) was added, the solution was mixed gently by inverting the tube several times, and incubated at room temperature for 5 min. 4 ml of pre-chilled QIAGEN buffer P3 (3.0 M potassium acetate, pH5.5) was added, the solution was mixed gently by inverting the tube several times, and incubated on ice for 15 min. The mixture was centrifuged at 20,000 x g (13,000 rpm for a Sorvall SS-34 rotor) for 30 min at 4°C, and the supernatant was carefully removed. Any suspended material would be re-centrifuged. QIAGEN-tip 100 column was equilibrated by applying 4 ml of buffer QBT (750 mM NaCl; 50 mM MOPS, pH7.0; 15% ethanol; 0.15% Triton X-100), after which the supernatant was applied to it. After all the solution entered the resin of the column, it was washed with 10 ml of QIAGEN buffer QC (1.0 M NaCl; 50 mM MOPS, pH7.0; 15% ethanol) twice, and DNA was eluted with 5 ml of buffer QF (1.25 M NaCl; 50 mM Tris–HCl, pH8.5; 15% ethanol). The DNA was precipitated with 0.7 volumes of room-temperature isopropanol and centrifuged immediately at 15,000 x g (11,000 rpm for a Sorvall SS-34 rotor) for 30 min at 4°C. The supernatant was carefully removed, the DNA pellet was washed with 70% ethanol, air-dried for 10 min, and redissolved in a suitable volume of TE buffer or nuclease-free water, depending on further applications.

Small-scale plasmid preparation.

For small-scale preparations Wizard® *Plus* SV Minipreps DNA Purification System (Promega) was used, following the manufacturer's protocol. This includes small scale method based on Birnboim protocol [1979], followed by the addition of a chaotropic salt (4.09 M guanidine hydrochloride) during neutralisation step. The cleared lysate after neutralisation and precipitation was applied to Wizard® *Plus* SV Minipreps Spin Column, the column was washed twice with Wizard® *Plus* SV Minipreps Column Wash Solution (162.8 mM potassium acetate, 27.1 mM Tris–HCl, pH7.5), and the DNA was eluted by nuclease-free water. The resulting plasmid DNA eluted from a column would normally be ethanol precipitated, unless it was used for restriction digestion only.

2.4.2. Phenol extraction and ethanol precipitation of DNA.

Phenol extraction followed by ethanol precipitation was the standard method of purification of crude DNA preparations and inactivation of enzymes after reactions.

An equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added to the DNA solution, the suspension was mixed vigorously by inverting the tube or by vortexing until homogeneity, and centrifuged in a microcentrifuge at 10,000 rpm for 1 min. The aqueous phase was carefully transferred to a clean Eppendorf tube. If a white precipitate was present at the aqueous/organic interface, the procedure was repeated (several times if necessary). This was followed by chloroform treatment, by adding an equal volume of chloroform and proceeding as above.

DNA was routinely precipitated from aqueous solution by adding 1/10 volume of 3M sodium acetate (pH5.5) and 3 volumes of absolute ethanol, mixing thoroughly and leaving on ice for a minimum of 10 min. This was then centrifuged in a microcentrifuge at maximal speed for at least 10 min at 4°C. The supernatant was discarded, the pellet washed with the same volume of 70% ethanol,

recentrifuged as above for 5 min and the supernatant was removed by aspiration. The pellet was left to dry on a bench for a couple of min. The dried DNA pellet could then be resuspended in a suitable volume of TE buffer (for storage or restriction digests) or water (for other manipulations). Sometimes, to keep the total volume small, 0.6 volume of isopropanol was used instead of ethanol. After isopropanol precipitation and centrifugation, the pellet was washed with 70% ethanol as described above.

2.4.3. Determination of DNA concentrations.

DNA concentrations were determined by measuring the absorption of appropriately diluted (usually 1:500) solutions at 260 nm, as an A^{260} value of 1.0 corresponds to double-stranded DNA concentration of 50 mg ml^{-1}. DNA purity was determined by measuring the ratio of absorption at 260 and 280 nm. Pure DNA should have a 260/280 ratio of at least 1.8. DNA concentration was always confirmed by running an aliquot on an agarose gel along with other DNAs of known concentrations.

2.4.4. Restriction of DNA.

Endonuclease reactions were typically performed in volumes of between 20 and 100 µl. The reactions contained the appropriate Boehringer Mannheim, New England Biolabs or Promega restriction buffer at 1x concentration. BSA was added (to 1 mg ml^{1}) if recommended by the manufacturer of an enzyme. The restriction enzyme was usually present in a two- to fivefold excess, i.e. 2–5 units per microgram of DNA. The digests were made up to their final volume using distilled water. The complete restriction mixtures were incubated at the temperature recommended for a particular enzyme for 1–3 h. The products of the reaction were either directly analysed by agarose gel electrophoresis, or phenol extracted, ethanol precipitated and dissolved in a suitable volume of TE buffer for further manipulations.

Partial digestion of DNA. For partial digestion of DNA, ten two-fold serial dilutions of restriction enzyme were added to fixed amounts of DNA, with 0.5 units of enzyme per µg DNA representing the highest enzyme:DNA ratio. The digests were incubated at the appropriate temperature for 1 h and terminated by addition of tracking dye. The products of the reactions could then be analysed by agarose gel electrophoresis.

2.4.5. Ligation of DNA.

DNA ligation was performed in a final volume of 10-20 µl. These usually contained between 0.5–1 µg of total DNA with insert DNA in a 2- to 20-fold molar excess over the vector DNA, 1x ligation buffer and T4 DNA ligase (Boehringer Mannheim or New England Biolabs). 0.2 unit (Weiss unit) of ligase was used for the ligation of cohesive DNA termini, and 1 unit of the enzyme for the ligation of blunt-ended molecules. The reactions were incubated for at least 30 min at room temperature for cohesive termini or at least 8 h at 16°C for blunt end ligation. Between 2 and 5 µl of the reaction mixture was directly used to transform competent cells of an appropriate strain of *E. coli.*

2.4.6. 'Filling in' of recessed 3' termini.

Sequenase® Version 2.0 T7 DNA polymerase (USB) was used to give blunt-ended DNA molecules by filling in the recessed 3' termini generated by various restriction enzymes. Reactions were performed in a final volume of 20 µl containing up to 1 µg DNA, 1x Sequencing buffer (USB), all four dNTPs each at a concentration of 20 mM and 2 units of the enzyme. The reactions were incubated at room temperature for 30 min. The reactions were stopped and the unincorporated nucleotides removed by increasing the reaction volume to 100 µl with TE, phenol extracting and ethanol precipitating the DNA.

2.4.7. Agarose gel electrophoresis.

Agarose gels for electrophoretic analysis of DNA were prepared with TAE buffer. The gels were made-up by melting the appropriate amount of agarose (usually between 0.8 and 1.5%) in 1x TAE buffer. Samples containing 1 x loading dye (6 x stock is 0.25% bromophenol blue, 0.25% xylene cyanol and 40% (w/v) sucrose in H_2O) were always loaded after immersion of the gels in TAE buffer. The gels were run at 80-100 mA at room temperature. Gels either contained 0.5 $\mu g\ ml^{-1}$ of ethidium bromide or were stained in water containing 2 $\mu g\ ml^{-1}$ ethidium bromide for about 20 min after completion of electrophoresis, and subsequently destained in fresh water for 5–10 min. After the run the DNA fragments were visualised by UV illumination and photographed.

2.4.8. Isolation of DNA from agarose gel slices.

DNA fragments of up to 15kb were isolated from agarose gels using Geneclean II kit from BIO 101. The procedure is based on a specific DNA binding to silica matrix (Glassmilk) at high salt concentration and its further release in low salt solution. The appropriate DNA bands were excised from ethidium bromide stained agarose gels under UV light in the minimal gel volume possible. The gel piece was placed in an Eppendorf tube, weighted and three volumes of 6 M NaI solution added. The tube was incubated at 50°C with periodic mixing until the agarose was dissolved (usually 2-5 min). 5 µl of Glassmilk was added, mixed and the tube stored on ice for at least 5 min with periodic mixing to keep the Glassmilk suspended. Silica matrix with bound DNA was pelleted by brief (5 seconds) centrifugation. The supernatant was discarded and the pellet washed three times with 400 µl of New Wash (an ethanol based buffer supplied with the kit), centrifuging and resuspending the pellet each time. The washed pellet was resuspended in 5 µl of TE buffer and incubated at 50°C for 2–3 min. The mixture was centrifuged for 30 seconds and DNA containing supernatant collected. The elution step was sometimes repeated to extract the DNA still bound to the matrix after the first elution.

2.4.9. Labelling DNA fragments by random priming.

The labelling of DNA fragments was performed using High Prime premixed solution (Boehringer Mannheim). The method is based on the hybridisation of nucleotides of all possible sequences to the denatured DNA to be labelled. High Prime mixture contains random oligonucleotides, Klenow polymerase, dATP, dGTP, dTTP and an optimised reaction buffer for rapid and efficient labelling of DNA with radioactive dCTP. 10 ng λ DNA and 100 ng of a plasmid fragment in 11 µl of H_2O were denatured by heating in a boiling water bath for 10 min and chilled quickly on ice. 4 µl of High Prime solution and 5 µl (50 µCi) [a32P]dCTP, 3000 Ci $mMol^{-1}$, was added to the DNA, and the solution was incubated for 10 min at 37°C. The reaction was stopped by addition of 2 µl 0.2 M EDTA. The labelled DNA was purified from low molecular size products on MicroSpin™ Column S-200 (Pharmacia).

2.4.10. Preparation of chromosomal DNA.

A single colony of the appropriate strain of *E. coli* was used to inoculate 5 ml of L-broth, which was incubated overnight at 37°C, (or 30°C for temperature-sensitive strains) with vigorous shaking. One millilitre of this culture was used to inoculate 100 ml of L-broth, which was again incubated at a suitable temperature overnight with constant agitation. The culture was then chilled on ice and transferred to a 250 ml centrifuge bottle and centrifuged at 5000 rpm. for 15 min at 4°C using a Sorvall GSA rotor. The supernatant was removed and the bacterial pellet was resuspended in 20 ml of STE, which is TE buffer with 10 mM sodium chloride. One millilitre of 10% SDS solution and 1 ml of proteinase K solution (4 mg ml^{-1}) were added, mixed gently and incubated at 50°C for 6 hours without shaking. To this solution an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added, the suspension mixed gently by inversion and allowed to stand at room temperature for 10 min. This was centrifuged in a benchtop centrifuge for 15 min to separate the aqueous and phenolic phases. The upper aqueous phase was then carefully removed avoiding the

protein interface. The nucleic acids in this phase were precipitated by adjusting the sample to 0.2 M sodium acetate (pH 5.5) and gently layering 2 volumes of ice-cold ethanol on top. The DNA was collected at the aqueous–ethanol interface by spooling it out with a glass rod.

The spooled DNA was washed in 70% ethanol, dried briefly in air and dissolved overnight in 10 ml of TE at room temperature. To this solution 0.1 ml of RNase A (10 mg ml^{-1}) was added and the mixture incubated at 37°C for 1 hour. Five hundred microlitres of a 10% SDS solution and 250 µl of proteinase K solution (4 mg ml^{-1}) were now added and the mixture incubated at 50°C for 1 h. The sample was extracted with phenol:chloroform:isoamyl alcohol and the DNA precipitated by spooling as above. After washing in 70% ethanol the DNA was air dried and solubilised in 1 ml of TE. This solubilisation took between 1 and 3 days. The yield of DNA was determined by UV spectrophotometry as described previously; from 100 ml of culture about 500 µg of chromosomal DNA was typically obtained.

2.4.11. Preparation of bacteriophage M13 DNA.

In all preparations of M13 DNA the *E. coli* strain TG1 was used. A 5 ml overnight culture of TG1 was prepared in the usual way. This culture was used to set up a culture infected with bacteriophage M13. Fifty microlitres of the overnight culture was used to inoculate 2 ml of L-broth. To this either 100 µl of an M13 phage suspension, (about 1/10 of a single plaque), or an entire M13 plaque from an agar plate was added. This culture was incubated at 37°C with vigorous shaking for about 5 h. One-and-a-half millilitres of this culture was transferred to an Eppendorf tube and centrifuged in a microfuge for 5 min. The resulting bacterial pellet could be used to prepare the double-stranded replicative form of M13 DNA, and the supernatant used to prepare single-stranded M13 DNA (or as a fresh bacteriophage suspension).

Preparation of double-stranded M13 DNA. The bacterial pellet was washed once in bacterial buffer and the double-stranded DNA

isolated in essentially the same way as was described for the small-scale isolation of plasmid DNA.

Preparation of single-stranded M13 DNA. 1.2–1.3 ml of the bacteriophage suspension was transferred to an Eppendorf tube and 200 µl of a solution containing 20% polyethylene glycol (PEG8000) in 2.5 M sodium chloride was added, the solution mixed thoroughly and allowed to stand at room temperature for 15 min. The precipitated bacteriophage particles were recovered by centrifuging in a microfuge for 5 min and the supernatant was removed carefully leaving the pellet as dry as possible. The bacteriophage pellet was resuspended in 100 µl TE with vigorous vortexing. Fifty microlitres of phenol (equilibrated with Tris–HCl (pH 8.0)) was added and the suspension mixed thoroughly by vortexing for 1 min. This was then centrifuged for 2 min and the upper aqueous layer carefully removed and placed in a fresh tube. The volume of the sample was adjusted to 0.5 ml and the single-stranded DNA recovered by phenol–chloroform extraction and ethanol precipitation as has been described earlier. An optional step was to chloroform extract the sample after the phenol–chloroform extraction. Using this method the yield of single-stranded DNA tended to be approximately 5–10 µg DNA per millilitre of infected culture and was of sufficient quality for both DNA sequencing and site-directed mutagenesis reactions.

2.4.12. Southern blotting procedures.

Transfer of DNA from agarose gels to a nylon membrane. The DNA was digested with the desired restriction enzymes, electrophoresed through agarose, stained and photographed as has been described above. The DNA fragments were depurinated by soaking the gel in 2 volumes of 0.25 M HCl with gentle agitation for 15 min. The gel was then rinsed with distilled water and the DNA denatured by soaking in 2 volumes of 0.5 M NaOH/1.5 M NaCl with gentle agitation for 30 min. The gel the was rinsed again in distilled water and placed in neutralisation buffer (1.5 M NaCl, 0.5 M Tris–HCl pH7.2, 1 mM

EDTA) for 15 min at room temperature with shaking. The last step was repeated.

A sheet of Hybond™-N+ membrane (Amersham) cut to the size of the gel saturated with 10 x SSC was placed on top of three sheets of blotting paper (saturated with the same buffer) on a clean glass plate. The pretreated gel was placed on top of the nitrocellulose, taking care to exclude any air bubbles; this was followed by a further three sheets of saturated blotting paper, and by another glass plate. The whole 'sandwich' was then inverted so that the glass plate which formed the base could be removed and a wad of dry paper towels placed on top. The whole structure was inverted again and weighted to provide even pressure and maintain good contact between the gel, nylon membrane, and paper towels. Transfer was allowed to continue overnight. After transfer was complete the membrane was removed, rinsed in 6x SSC for 2 min, blotted dry, and the DNA was fixed by UV crosslinking.

Pre-hybridisation of membranes. The nylon membrane was pre-hybridised for 1–2 hours at 65°C in 20 ml of pre-hybridisation solution (typically 0.1 ml solution per cm^2 of filter) in a hybridisation oven.

Hybridisation of membranes. A labelled DNA probe was boiled for 5 min and cooled on ice before adding to the hybridisation tube into pre-hybridisation solution, and hybridisation was carried out at the same temperature overnight.

The membrane was rinsed briefly at room temperature in Wash buffer. The membrane was then washed 3 times in 250 ml of Wash buffer for 10–15 min at 30°C. Membranes are then usually ready for exposure and detecting the presence of the probe DNA either by autoradiography, or by phosphoimaging analysis.

Solutions for Southern hybridisation.

20 x SSC

3 M NaCl
0.3 M sodium citrate

Pre-hybridisation and hybridisation solution

7% SDS (7 ml of 20% stock solution)
0.5 M sodium phosphate buffer pH 7.2
1 mM EDTA

Wash buffer

0.5% SDS
0.5 M sodium phosphate buffer pH 7.2
1 mM EDTA

2.4.13. Amplification of DNA by polymerase chain reaction

Polymerase chain reaction (PCR) was used for amplifying fragments of known DNA. Plasmid or chromosomal DNA was used as template. Oligonucleotide primers were obtained from various commercial sources. Vent DNA polymerase from New England Biolabs was used when higher fidelity of reaction was required, otherwise *Taq* polymerase from Promega was used. Buffer supplied by the manufacturer was used with each polymerase. Magnesium was usually present in Taq buffer. When using Vent polymerase, magnesium chloride was added to concentration of 2 mM. The usual reaction mix was as follows:

Polymerase buffer (10x)	5 µl
dNTP mix (5 mM each)	2 µl
Oligonucleotide primers	10 pmol each
Thermostable DNA polymerase	1 unit
distilled water	to 50 µl

Reactions were performed in a Techne thermal cycler in a volume of 10–100 µl with a layer mineral oil on top of the reaction mix. Typical

cycling parameters were denaturing at 94°C (95°C for Vent polymerase) for 1 min, annealing for 1 min at the appropriate temperature for the primer-template pairs (usually 5–10° lower than theoretical T_m of the primers) and polymerisation at 72°C (75°C for Vent polymerase) for 2 min (plus 1 min for every kb of template above 1,000 bases).

2.4.14. Site-directed mutagenesis.

Mutagenesis based on PCR.

The described technique was used to mutagenise several bases in *rpoD* which were lying closely enough to *Bam*H I site within this gene to be covered by one oligonucleotide, therefore the mutagenesis can be carried out in one step. A mutagenising primer had the *Bam*H I site near its 5' terminus. The second primer for PCR was about 300 bp away on the template DNA, and between it and the first, mutagenising, primer there was a *Xho* I site.

To optimise PCR the template concentration was increased approximately 100-fold over conventional PCR conditions (approximately 0.5 pM), the number of cycles was reduced from 25–30 to 5–10, and a high-fidelity $Vent_R$ DNA polymerase (New England Biolabs) was used. This steps would prevent clonal expansions of any undesired second-site mutations.

After the completion of PCR and checking an aliquot on an agarose gel electrophoresis, the DNA was precipitated, digested with *Xho* I and *Bam*H I and cloned into a plasmid (pTK111) which allowed for an enrichment of the ligation product. The ligation mixture was then transformed into DH5α competent cells, plasmid DNAs from three transformants were purified, and the region between *Xho* I and *Bam*H I sites was sequenced. Unexpectedly, given a relatively short length of PCRed and cloned DNA and all the PCR optimising conditions, about a half of all the sequenced inserts had a second undesired mutation. The plasmids with the desired DNA sequences were used for further manipulations.

Mutagenesis after Kunkel.

E. coli strain BW313 which is *dut-1* and *ung-1* and therefore allows incorporation of uracils into phage M13 DNA was transformed with replicative form of M13mp18–14 DNA. A single plaque was picked up and propagated for approximately 5 hours at 37°C in 5 ml L-broth supplemented with 25 mg l^{-1} uridine.

After this the relative titres of the bacteriophage suspension were determined on strains BW313 *(dut-1 ung-1)* and TG1 *(dut-1$^+$ ung-1$^+$).* M13 lysate was diluted down to 10^{-10} (initial 1:100 followed 1:10 dilutions). 100 µl of each dilution from 10^{-7} to 10^{-10} was mixed with 200 µl of overnight cultures of both strains, incubated for 10 min at 37°C, after which 3 ml of molten top agar was added and the mixture was plated on L-agar. If the titre of strain BW313 is less than five orders of magnitude than that of strain TG1, the phage particles would undergo another passage on BW313 cells.

When uracil enrichment which is the ratio between titres of the phage on BW313 and TG1 strains was better than 10^5 (after 5 passages) the uracil-enriched single stranded DNA template was isolated.

The mutagenesis was performed by annealing a 5'-phosphorylated mutagenising oligonucleotide to the U-rich single-stranded template DNA (at a molar ration 20:1) in 1 x Sequenase buffer (USB) (total volume 20 µl) by placing the tube into a beaker with water at 75°C (20° above the theoretical T_m for a primer) for 5 min and then cooling slowly in the same beaker to 30°C for approximately 40 min.

To synthesise the second strand the following mixture was added to the reaction:

$MgCl_2$ 200 mM	0.5 µl
DTT 0.1 M	0.5 µl
dNTPs 2.5 mM	4 µl
ATP 10 mM	2 µl
T4 DNA ligase (Boehringer Mannheim)	1 U
Sequenase T7 DNA polymerase (USB)	1U

The reaction was incubated on ice for 5 min, at 25°C for 5 min and at 37°C for 90 min. After this the reaction was placed on ice, 80 µl EDTA buffer (Tris–HCl 10 mM pH 8.0, EDTA 10 mM) was added to it, and 3–10 µl of it was used to transform 300 µl competent TG1 cells. Then 50–300 µl exponentially growing TG1 cells were added, and the phage was plated out. Single plaques were picked up, single stranded and RF phage DNA was isolated and the region which would be further cloned was sequenced. All the sequenced phage DNAs contained desired and no secondary mutations.

2.4.15. DNA sequencing.

DNA sequencing was performed using the Dye Terminator Cycle Sequencing kits with AmpliTaq® Polymerase (Perkin Elmer). Gel images were analysed with Sequence Navigator computer package.

2.5. Protein Techniques.

2.5.1. Polyacrylamide gel electrophoresis of proteins.

E. coli proteins were routinely separated using SDS–polyacrylamide gel electrophoresis with a discontinuous buffer system [Laemmli, 1970]. 7–12% resolving gel and 4% stacking gel were employed. The composition of typical gel solutions was as follows:

10% Resolving gel.

acrylamide stock solution (40%)		6.25 ml
4x resolving gel buffer		6.25 ml
glycerol 50%		5 ml
7.5% ammonium persulphate (freshly prepared)		0.25 ml
distilled water		12 ml
TEMED		15 µl
	Total	25 ml

4% stacking gel.

acrylamide stock solution (40%)		1.0 ml
4x stacking gel buffer		2.5 ml
7.5% ammonium persulphate (freshly prepared)		0.1 ml
distilled water		6.3 ml
TEMED		10 µl
	Total	10 ml

These solutions were made up immediately prior to use, with the TEMED being added last before casting a gel. The depth of the stacking gel between the bottom of the comb and the resolving gel was about 1 cm. Once the resolving gel had been poured the stacking-gel solution was carefully poured on top of the resolving gel, the comb inserted and polymerisation allowed to occur. The comb was then removed and the wells washed out with 1x reservoir buffer, which was also used to fill up the buffer chambers of the apparatus. The sample could be loaded onto the gel at this stage.

Samples were mixed 1:1 with 2x PAGE-loading buffer, kept at 37°C for 1 hour or boiled for 3 min if the whole cell lysates were analysed, and centrifuged in a microcentrifuge for 5 min prior to loading. Gels were typically electrophoresed until the bromophenol blue dye-front had run off the bottom of the gel. Once electrophoresis was complete the gel was removed from the apparatus and stained.

If stained with Coomassie, the gel was placed in staining solution for 45–60 min with constant gentle agitation. Gels were the transferred into destaining solution and left for anything between

2–24 h. For preservation, the stained/destained gel was soaked in destaining solution plus 5% glycerol for 30 min and dried down on blotting paper using a vacuum gel drier at 80°C for about 1 h.

If reverse-stained [Fernandez-Patron *et al.*, 1992], the gel was rinsed with water, incubated in 0.2 M imidazole, 0.1% SDS with agitation for 15–20 min, rinsed with water again and incubated with 0.2 M Zn_2SO_4 for 120 seconds exactly, after which the staining solution was rapidly diluted with water and the gel was rinsed.

Solutions used in SDS–PAGE (glycine SDS–PAGE).

4x stacking-gel buffer (0.5 M Tris)

15.25 g of Tris base, dissolved in 200 ml distilled water, adjusted to pH 6.8 with concentrated HCl; 0.4% (w/v) SDS, made up to 250 ml, filtered and autoclaved.

4x resolving-gel buffer (1.5 M Tris)

45.5 g of Tris base, dissolved in 200 ml distilled water, adjusted to pH 8.8 with concentrated HCl; 0.4% (w/v) SDS, made up to 250 ml, filtered and autoclaved.

10x reservoir buffer

30.2 g of Tris base, 144 g of glycine dissolved in 600 ml distilled water, made up to a final volume of 1 litre and filtered. SDS was added to 0.1% in the final 1x buffer.

2x PAGE loading buffer

4x stacking gel buffer		0.125 ml
10% SDS		0.300 ml
50% glycerol		0.200 ml
2–mercaptoethanol		0.050 ml
0.1% bromophenol blue		0.200 ml
Distilled water		0.125 ml
	Total	1.0 ml

Stock acrylamide

37 g acrylamide, 1 g NN'-methylene-bis–acrylamide, made up to 100 ml with distilled water, filtered and stored at 4°C.

Coomassie staining solution

9% (v/v) acetic acid, 45% (v/v) methanol, and 0.1% (w/v) Coomassie brilliant blue.

Destaining solution

7% (v/v) acetic acid, and 5% (v/v) methanol.

Tris-tricine SDS–PAGE.

Tricine allows a resolution of small proteins at lower acrylamide concentrations than in glycine-SDS-PAGE systems.[Schagger and von Jagow, 1987]. For each minigel (Hoeffer "Mighty Small" or BioRad "Mini Protean-II") the following solutions were prepared::

10% separating gel

1.22 ml separating acrylamide (46.5% acrylamide/1.5% bis-acrylamide)
2 ml gel buffer
2 ml 50% glycerol
0.78 ml water
75 µl 10% APS
7.5 µl TEMED

Stacking gel:

0.25 ml stacking acrylamide (48% acrylamide/1.5% bis-acrylamide)
0.75 ml gel buffer
2.0 ml water
20 µl 10% APS
2 µl TEMED

Both the stacking and separating gels were polymerised at the same time, and stacking gel was carefully poured directly onto the

separating gel. Each separating gel mixture was made separately and TEMED and APS were added right before pouring. Electrophoresis was carried out at 25–35 mA (constant current) per gel in minigel setup, and additional cathode buffer was added during the run. After electrophoresis was complete gels were stained as above.

Solutions used in tricine SDS–PAGE.

Anode buffer (+) (can be diluted from 10 x stock)

200 mM Tris pH 8.9

Cathode buffer (-) (can be diluted from 10 x stock)

100 mM Tris
100 mM Tricine
0.1% SDS (no need to pH, but it should be ~8.25)

Gel buffer

3.0 M Tris pH 8.45
0.3% SDS

The pH of the anode and gel buffers are essential.

2 x sample buffer, for 20 ml:

5 ml 0.5 M Tris, pH 6.8
4 ml 20% SDS
1 ml 2-mercaptoethanol
4 ml 50% glycerol
0.004 g bromophenol blue
6 ml water

2.5.2. Western blotting procedures

Western blotting was performed using a BioRad electrophoretic transfer cell. The previously run gel was trimmed down by removing the stacking gel and placed on top of several sheets of blotting paper saturated in protein-transfer buffer (which is 1x protein gel

reservoir buffer but without any SDS). A sheet of nitrocellulose saturated in the same buffer was placed on top of the gel excluding any bubbles followed by several more sheets of saturated blotting paper. The whole sandwich was secured in the transfer apparatus and transfer was allowed to continue at 30 mA for 14 hours.

After this, non-specific binding sites were blocked by immersing the membrane in 5% blocking reagent (skimmed milk) in PBS-T buffer for one hour. The membrane was briefly rinsed twice with fresh changes of PBS-T, then washed three times, once for 15 min and twice for 5 min, with fresh changes of washing buffer. All washing steps, as well as incubations with antibodies were performed at room temperature with gentle agitation. At least 200 ml of the buffer was used for each washing step.

The membrane was incubated in the appropriately diluted primary antibody (mouse monoclonal antibody to *E. coli* β-galactosidase, Boehringer Mannheim) for 1 hour. Dilutions ranged from 1:500 to 1:2000. The incubation was performed in square plastic dishes with 5 ml of antibody dilution. The washing step described above was repeated and the membrane was incubated in the secondary antibody (horseradish peroxidase conjugated anti-mouse Ig, Amersham) diluted 1:1000 for 1 hour. The filter was then removed and washed in PBS-T as before.

The blot was removed and treated with 1:1 mixture of Amersham "ECL" detection reagents, with just enough liquid to cover the blot, for 1 min. Then the blot, excess liquid removed, was wrapped in SaranWrap and quickly exposed to a film. A series of exposures was performed, usually from 15 seconds to several hours.

Phosphate buffered saline - Tween (PBS-T) pH 7.5

di-sodium hydrogen orthophosphate anhydrous	11.5 g
sodium dihydrogen orthophosphate	2.96 g
sodium chloride	5.84 g
Tween 20	0.1% (v/v)
distilled water	to 1 litre

2.5.3. Overexpression of proteins in *E. coli* cells.

Overexpression of $\sigma^{\Delta 245\text{-HMK Y425C}}$, $\sigma^{\Delta 245\text{-HMK S428C}}$, $\sigma^{\Delta 245\text{-HMK Q437C}}$, and $\sigma^{\Delta 245\text{-HMK K502C}}$ was carried out as followed. *E. coli* BL21(λDE3) cells were transformed with pTK413, pTK418, pTK420, and pTK422. Overnight culture of a purified transformant was diluted 1:100 in 1 l L-broth supplemented with glucose 13–15 g l^{-1} and ampicillin 100 mg l^{-1}. The culture was grown in 1 l fermenter with constant aeration, and in a mid-log phase at A^{600} of 3–4 T7 RNA polymerase was induced by addition of IPTG to final concentration 1.5 mM. pH of the medium was kept within 7.4–7.6 with automatic addition of concentrated HCl or 10 M NaOH solutions. Cells were harvested 3 hours after IPTG induction and were ready for isolation of inclusion bodies.

Overexpression of ScrP and MtgA was performed according to Khattar *et al.* [1994]. *E. coli* BL21(λDE3)pLysS cells were transformed with plasmids which have a gene encoding a protein to express under T7 ∅10 promoter. Cells were grown in minimal medium supplemented with glucose (0.4%), thiamine (1 μg ml^{-1}), ampicillin (100 μg ml^{-1}), and chloramphenicol (20 μg ml^{-1}) at 37°C until they reached A^{600} 0.7. Then 0.5 ml was removed from each culture, and the expression of the T7 RNA polymerase was induced by addition of IPTG to final concentration 0.5 mM. Incubation was continued for 30 min before rifampicin was added to final concentration 200 μg ml^{-1} host DNA polymerase. After 45 min, each of the samples was pulse-labelled with 5 μCi of [^{35}S]methionine (Amersham) for 5 min, after which cells were cooled on ice, harvested, and resuspended either in 200 μl of SDS–PAGE sample buffer, or subjected to further cell fractionation.

2.5.4. Isolation of inclusion bodies.

Cells were pelleted at 6000 x g at 4°C for 10 min. Cell pellet was homogenised in 160 ml of lysis buffer, the mixture was lysed by ultrasonication, stirred at 4°C for 30 min and centrifuged at 30,000 x g at 4°C for 20 min. The pellet was washed twice with 80

and 40 ml of NaDoc buffer, once with 40 ml of triton buffer, and centrifuged as above. The resulting pellet was dissolved in 20 ml of 6M guanidine chloride in TGEB buffer.

Lysis buffer:

50 mM Tris HCl pH8 (at 4°C)
1 mM EDTA
100 mM NaCl
0.25 mM PMSF (Phenylmethylsulphonyl fluoride)
(stock solution 0.1 M in *i*-propanol)
4 mg l^{-1} Lysosyme

NaDoc buffer:

50 mM Tris HCl pH8 (at 4°C)
1 mM EDTA
100 mM NaCl
0.5% sodium deoxycholate

Triton buffer:

50 mM Tris HCl pH8 (at 4°C)
1 mM EDTA
100 mM NaCl
0.5% Triton X-100
10 mM β-mercaptoethanol

TEB buffer:

50 mM Tris HCl pH8 (at 4°C)
1 mM EDTA
10 mM β-mercaptoethanol

TGEB buffer:

TEB buffer with 5% glycerol

2.5.5. Protein purification from inclusion bodies.

After washing and denaturing inclusion bodies a novel procedure [Shimamoto *et al.*, 1998] was applied for renaturing, in which a

protein solution containing 6 M guanidine chloride was diluted 2-fold with glycerol to prevent precipitation during the dialysis, placed into dialysis tubing and dialysed against 10 volumes of 1:3 (v/v) mixture of TEB and glycerol at 4°C during 10–16 hours and 3–4 changes of the outer solution during the dialysis. The dialysis was performed in 50 ml plastic tubes which were constantly rotated. The dialysis was monitored by the conductivity of the outer dialysis buffer. After the dialysis and centrifugation the protein solution was rapidly diluted against 10 volumes of 0.3 M ammonium sulphate in TGEB buffer, and the proteins were precipitated at 4°C by adding ammonium sulphate to 1 M.

The proteins were dissolved in 6 M guanidine chloride/TGEB buffer. Phenyl-Toyopearl column with the bed volume 11 ml was equilibrated with 6 M guanidine chloride 0.9 M ammonium sulphate in TGEB buffer. Protein solution in 6 M guanidine chloride/TGEB buffer were loaded on the column and eluted with a 0.9 to 0 M gradient of ammonium sulphate in 6 M guanidine chloride/TGEB buffer.

Fractions containing the least impurities were pooled, diluted 2-fold with glycerol to prevent precipitation during the dialysis, and dialysed as before against 10 volumes of 1:3 (v/v) mixture of TEB and glycerol. The proteins were stored at –20°C and rapidly diluted in reaction buffer before usage.

2.5.6. Column chromatography.

Chromatography was performed on ion exchange columns POROS HQ/M, POROS SP/M (PerSeptive Biosystems), with 0.2 to 1 M NaCl gradient in TGEB buffer with 0.025% Tween-20, at flow rate 0.5 ml min^{-1}, total back pressure 1.2–1.4 mPa, using SMART system (Pharmacia).

Chromatography on Superose 6 column (Pharmacia) was performed at 10°C in 0.2 M NaCl, 0.025% Tween-20 in TGEB, at flow rate

40 µl min^{-1}, total back pressure 1.2–1.4 mPa, using SMART system (Pharmacia).

Chromatography on columns filled with hydrophobic resin Phenyl Toyopearl M (Tosoh, Tokyo) was performed under denaturing conditions using an HPLC system. Phenyl-Toyopearl column with the bed volume 11 ml was equilibrated with 6 M guanidine chloride 0.9 M ammonium sulphate in TGEB buffer. Protein solution in 6 M guanidine chloride/TGEB buffer were loaded on the column and eluted with a 0.9 to 0 M gradient of ammonium sulphate in 6 M guanidine chloride/TGEB buffer.

2.5.7. Protein concentration determination.

Concentrations of proteins were determined by BCA protein assay (Pierce, Rockford). Calibration curves were constructed using 0–2 mg ml^{-1} BSA in "outer buffer solutions", i. e. solutions against which σ subunits were dialysed at the final of the purification process. Series of reactions were performed at room temperature for up to 60 min in microtitre plates and absorbance at 562 nm was measured.

2.5.8. Preparation of total and outer *E. coli* membranes for protein analysis [K. Begg, pers. comm].

The bacterial cells (0.5 ml) were harvested by centrifugation (3600 g, 10 min at 4°C) and the pellet was resuspended in 0.5 ml of a solution containing 10 mM Tris–HCl pH 7.6, 5 mM EDTA, 250 µM PMSF and 2 mM 2-mercaptoethanol. The cells were then lysed by sonicating four times for 15 sec each at a setting of 5–6 on the MSE sonicator with constant chilling on ice. Unlysed cells and other debris were removed by centrifuging as above and the resulting supernatant ultracentrifuged at 68,000 rpm. in a Beckman TLA100.2 rotor for 1 h at 4°C. The membrane pellet was resuspended in 2 ml of the above buffer and ultracentrifuged as before. The pellet was then resuspended in 40 µl of the same buffer and was then ready for SDS–PAGE analysis of the resultant total membrane proteins.

For preparation of outer membrane proteins the above procedure is followed up to pelleting membranes for the first time. This pellet is resuspended in 0.5% Sarcosyl containing 250 μM PMSF by passing gently through a 23 gauge needle. The outer membrane is then pelleted by centrifuging for 1 hour at 100,000 x g. This pellet is resuspended again in 0.5% Sarcosyl/250 μM PMSF and centrifuged. This pellet was resuspended directly into the loading buffer for SDS–PAGE.

2.5.9. Preparation of periplasmic protein fraction.

The bacterial cells (0.5 ml) were harvested by centrifugation (8000 x g, 10 min at 4°C) and the pellet was resuspended in 125 μl of a solution containing 20% sucrose (w/v), 50 mM Tris–HCl (pH 7.4), 250 μM PMSF and 1 mM EDTA. The suspension was incubated on ice with shaking for 10 min after which the cells were harvested as above. The supernatant was removed and the thoroughly drained cell pellet resuspended, this time in 50 μl of ice-cold 5 mM $MgCl_2$/250 μM PMSF. Again the suspension was incubated on ice with shaking for 10 min and clarified by centrifugation.

The remaining supernatant (termed the cold osmotic shock fluid) was then analysed on SDS–PAGE. The pellet was washed again with 5 mM $MgCl_2$/250 μM PMSF, re-precipitated, dissolved in the buffer for total membrane isolation. After precipitation of total membrane fraction the supernatant was concentrated by acetone precipitation (with 5 volumes of acetone) and resuspended in the loading buffer for SDS–PAGE. This solution was used as a "cytoplasmic fraction" on some protein gels.

2.5.10. Hybrid σ assay [Kumar *et al.*, 1995a].

A hybrid gene encoding the amino-proximal 529 aa (or its derivatives) of the major *E. coli* σ subunit, $\sigma^{70/32}$ (including region 2.4) followed by the last 82 aa of the heat-shock σ protein, σ^{32} (including region 4.2) is present under the control of P_{lac} on pHW111 (or its derivatives). This plasmid also has a Kan^R selection

marker. *lacI*^{Q}, and p15A replicon. A reporter gene, *lacZ*, is present on the second plasmid, pML$_{HM}$::*lacZ*, under control of a hybrid promoter with regions –35 for σ^{32} and –10 for σ^{70}. pML$_{HM}$::*lacZ* has pMB1 replicon and Amp^{R} gene (Fig. 2.1).

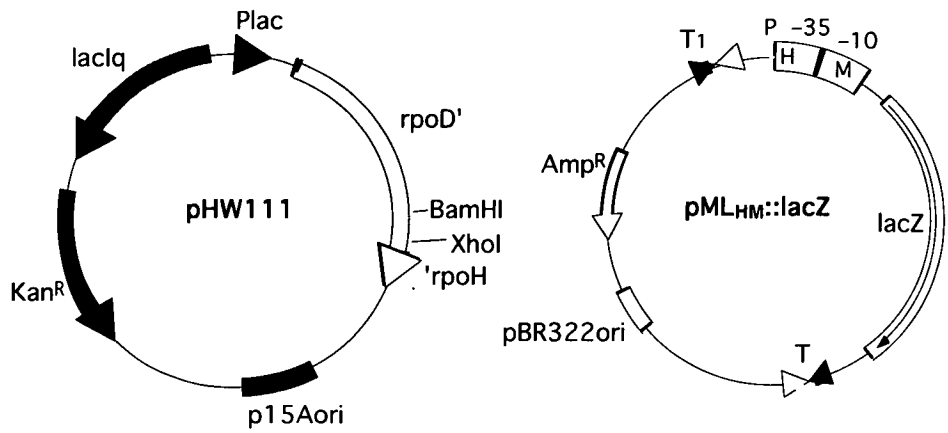


Fig. 2.1. Two compatible plasmids, hybrid σ donor pHW111 and reporter pML$_{HM}$::*lacZ*, used in the hybrid σ assay. T and T1 are transcription terminators in both directions. H and M are –35 and –10 promoter elements for σ^{70} and σ^{32}, respectively.

NM570 strain carrying both hybrid σ and reporter plasmids were grown in Spitzizen minimal glucose medium supplemented with thiamine, Amp and Kan, at 37°C. The cultures were then diluted to A^{600} 0.08 with fresh warm medium and grown to A^{600} 0.6–0.7 in the presence or absence of IPTG (15 μg ml^{-1}) at a required temperature. At this point samples were taken and assayed for β-galactosidase activity, as described in the next section. The activity of a hybrid σ was determined as difference between activities of IPTG-induced and non-induced samples. These were compared against the activity of a "wild-type" hybrid σ encoded by pHW111.

2.5.11. β-galactosidase activity assay.

The method used was essentially as described by Miller [1982]. Samples of either 0.5 ml or 0.1 ml (depending on β-galactosidase activity of the strain) were taken and added to 0.5 ml or 0.9 ml of Z-buffer, respectively. To permeabilise the cells Z-buffer contained 0.005% SDS and in addition 15 μl of chloroform was added to each

sample. The sample was vortexed for 10 seconds and could at this point be sealed and stored overnight at 4°C.

To assay, 200 µl of 4 mg ml^{-1} ONPG (*o*-nitrophenyl-β-D-galactoside, Sigma) in 0.1 M MOPS (pH7.0) was added to each sample. The samples were incubated at 30°C until sufficient yellow colour (*o*-nitrophenol) had appeared such that when 0.5 ml of 1 M Na_2CO_3 was added to stop the reaction, the absorbance at 420 nm was between 0.1 and 2.0 (absorbance measurements are linear within this range). Samples were spun in a microcentrifuge to remove cell debris. The absorbance at 420 nm was determined.

Total enzyme activity was calculated as: $1000 \times A^{420} / (t \times V)$, where t is incubation time in minutes and V is the volume of the samples in millilitres (generally 0.5 or 0.1).

The specific activity is defined as total enzyme activity divided by A^{600} of the culture at the time of sampling (i. e. enzyme units per total cell mass) and is expressed in Miller Units (MU).

Z-buffer: 60 mM Na_2HPO_4
40 mM NaH_2PO_4
1 mM $MgSO_4$
10 mM KCl
0.005% SDS
0.05 M β-mercaptoethanol (3.51 ml per 1l)

2.5.12. Selection for *kat* (RpoS$^-$) phenotype.

rpoS mutants have reduced expression of katalase activity in stationary phase as a result of decreased expression of σ^s-dependent *katE* gene. The production of oxygen upon addition of hydrogen peroxide was tested by adding 10 µl 3% H_2O_2 to 50 µl of culture that had been grown overnight in L-broth. Bubbling (oxygen elevation) was exhibited by *rpoS*$^+$ cells immediately upon the addition of the agent, but rpoS::KanR strain exhibit reduced/delayed activity [Farewell *et al.*, 1996].

2.6. RNA polymerase techniques.

Molar concentrations of RNA polymerase holoenzymes are concentrations of their core enzymes, the molarity of a σ subunit shown as a proportion to the core.

2.6.1. Promoter DNA templates for transcription.

Preparation of T7A1 promoter DNA template was carried out as described before [Dunn and Studier, 1983] by PCR a fragment using pHR1435 as a template. For the gel shift assays, one of the primers was end-labelled with γ-[32P]ATP. The product was subjected to 8% native PAGE and stained with ethidium bromide. The band of 180 bp was cut out of the gel, incubated overnight in 5 volumes of the running buffer at 4°C, and the DNA was extracted using QIAquick DNA purification columns. The concentration of the DNA was estimated by comparing the densities of ethidium bromide stained bands in agarose gel. with those of DNA standards.

The template carrying λP_R promoter having 32 bp initial A-less transcribed sequence instead of the original sequence was prepared by PCR of a 190 bp DNA fragment from AL1130 template [Sen *et al.*, 1998; Kubori and Shimamoto, 1996]. The distance between the transcriptional start and the upstream end of the fragment is 85 bp, with the overlapping λP_{RM} promoter being inactivated.

2.6.2. *In vitro* transcription.

Before carrying out transcription reactions, holoenzymes were formed by incubation of the core enzyme with a σ subunit at ratio 1:4 to 1:10 (as was earlier determined for every holoenzyme) in TT buffer at 37°C for 10 min. Before adding substrates, RNA polymerase and DNA were pre-incubated for 10 min at 37°C in TT buffer. Transcription was carried out in 8 µl of TT buffer by adding 5 µM 100 Ci mmol^{-1} of the initiating nucleotide [γ-^{32}P]GTP, 100 µM UTP and CTP for λP_R promoter, or 5 µM 100–400 Ci mmol^{-1} [γ-^{32}P]ATP and 100 µM of each of the other NTPs for T7A1 promoter,

for 20 min. Heparin (40 μg ml^{-1}) was added together with the substrates to prevent reassociation of released RNA polymerase with DNA.

For pulse-labelling experiments, the unlabelled substrates (5 μM GTP, 100 μM UTP and CTP) and 40 μg ml^{-1} of heparin were first added, and then after the indicated time the reaction was chased with 5 μM 100 Ci $mmol^{-1}$ [γ-^{32}P]GTP for 5 min.

Reactions were stopped by mixing with the equal volume of phenol/chloroform/isoamyl alcohol (25:24:1), the transcripts were separated on 20% polyacrylamide–7M urea gel. The gels were analysed and radioactivity was quantified on a BAS2000 Image Analyser (Fuji Film, Tokyo).

TT buffer

50 mM Tris–HCl, pH7.9
100 mM KCl
10 mM $MgCl_2$
1 mM DTT
0.1 mg ml^{-1} partially hydrolysed casein
0.025% Tween 20

Gel electrophoresis of run-off transcripts.

A DNA sequencing apparatus was used for separating RNA transcripts. The glass sequencing gel plates were thoroughly cleaned with detergent and ethanol, assembled using 0.2 mm spacers and taped together to minimise the possibility of leakage.

The gel was prepared by adding together the following:

40 % (w/v) acrylamide solution*	50 ml
urea	43 g
10x TBE	10 ml
water	to 100 ml
ammonium persulphate 10% w/v	1 ml
TEMED	35 μl

*19:1 mix of acrylamide and N,N`-methylenebisacrylamide

This solution was poured between the sequencing plates. After polymerisation the gel was then clamped into the sequencing apparatus and 1x TBE solution poured into the top and bottom reservoirs. The gel was then pre-run at about 50 W (~1500 V) for 1 h and then loaded with the sequencing reactions. The gel was then electrophoresed at 50 W until the blue dye-front ran off the end of the gel. Once electrophoresis was complete the top plate was removed and the gel was transferred to a sheet of wet blotting paper, covered with SaranWrap and dried in a vacuum gel-drier for 1 h at 80°C. The gel was then analysed and radioactivity was quantified on a BAS2000 Image Analyser (Fuji Film, Tokyo).

2.6.3. Gel-shift assay.

Binary complexes were formed by incubation of RNA polymerase with 10 or 100 nM DNA in 10 µl TT buffer in the presence of 10% glycerol at 37°C for 10 min. Samples were then loaded onto a running 4% native polyacrylamide gel containing 45 mM Tris-borate, pH 8.0, and 1 mM EDTA. The dye (30% glycerol, 0.2% bromophenol blue, 0.2% xylene cyanol) was loaded in a separate lane. If heparin was used, 0.5 µl of 1 mg ml^{-1} heparin (to 50 µg ml^{-1}) was added 15 seconds prior to loading. The electrophoresis was carried out at 20 mA at room temperature. The gels were analysed and radioactivity was quantified on a BAS2000 Image Analyser (Fuji Film, Tokyo).

2.7. Sequence analysis.

Most sequences were obtained from the EMBL, GenBank, SwissProt or TrEMBL databases. Sequence similarity searches were performed with *BLAST2* programmes [Altschul *et al.*, 1997] using usually network server at NCBI. Preliminary sequence data was obtained from The Institute for Genomic Research website at http://www.tigr.org. Sequences from the unfinished genome projects (as of September, 1999) were searched using the *BLAST2* programme at www.ncbi.nlm.nih.gov/BLAST/unfinishedgenome.

html; these sequences however still contained errors. The *ClustalW* programme [Thompson *et al.*, 1994] available at ClustalW WWW Service at the European Bioinformatics Institute (www2.ebi.ac.uk/clustalw) was used for multiple sequence alignment.

Prediction of signal peptides of the proteins and their localisation was performed using *SignalP 1.1* [Nielsen *et al.*, 1997] (at www.cbs.dtu.dk/services/SignalP/) and *PSORT* [Nagai, 1997–1999] (at http://psort.nibb.ac.jp:8800/) programmes, the latter also uses other criteria than the presence of signal peptide to predict a peptide localisation.

Prediction of transmembrane regions in polypeptides was performed using programmes linked to the ExPASy Molecular Biology Server. These include *PSORT* (see above),

DAS [Cserzo *et al.*, 1997]
at www.biokemi.su.se/~server/DAS/,

TMHMM 1.0 [Sonnhammer *et al.*, 1998]
at www.cbs.dtu.dk/services/TMHMM-1.0/,

TMpred [Hofmann and Stoffel, 1993]
at www.ch.embnet.org/software/TMPRED_form.html,

PHDhtm [Rost *et al.*, 1995] at the Predict Protein server
http://dodo.cpmc.columbia.edu/predictprotein/,

HMMTOP [Tusnady and Simon, 1998]
at www.enzim.hu/hmmtop/, and

SOSUI [Hirokawa *et al.*, 1998]
at http://azusa.proteome.bio.tuat.ac.jp/sosui/.

Hydrophobicity profiles were built with the *ProtScale* programme at the ExPASy Molecular Biology Server utilising Kyte-Doolittle [Kyte and Doolittle, 1982] hydrophobicity scales.

2.8. Image analysis and densitometry.

Quantitative results were obtained by comparing sample band intensities on the scanned blot images with samples of known

concentrations of a standard, using the NIH image 1.62b7 programme (http://rsb.info.nih.gov/nih-image). However, many protein measurements were not quantitative, and only relative amounts of proteins were measured.

Analysis of gels carrying ^{32}P or ^{35}S label was usually performed on a BAS2000 Image Analyser (Fuji Film, Tokyo) or on a PhosphoImager (Molecular Dynamics).

Chapter 3

Studies on structure/function relationship in the σ^{70} subunit of RNA polymerase

3.1. Substitution of single *rpoD* codons for amino acid residues in regions 2.3, 2.4, 3.1 and 3.2 of σ^{70} with cysteine.

For this study several amino acid residues were chosen to be substituted with cysteines: Y421C, Y425C, Y425C W433A, S428C, Q437C, T440C, and S442C. These amino acids had been demonstrated to participate in promoter recognition (in the σ^{70} conserved region 2.4) or were suspected to be involved in promoter melting (region 2.3). Besides, substitutions to cysteine in these cases were thought unlikely to cause loss of σ^{70} activity, as was described in Chapter 1. If this is the case and these substitutions result in functional σ^{70} proteins, then they can be used to probe the environments around these cysteines and study the contacts made with surrounding DNA as well as peptides, at different steps of promoter opening.

The non-conserved K502 residue in region 3.1 was also chosen for study, as it lies close to the conserved P504 and S506 residues which play a role in abortive synthesis (Chapter 1). Similarly, the less conserved S517 residue in region 3.2 was chosen, as it was thought likely that it lies close to the transcriptional start, as confirmed later by Owens *et al.* (see Chapter 1).

Several vectors were prepared as templates for site-directed mutagenesis. Firstly pTK210 and pTK310 were constructed. pTK210 is a derivative of pGEMDΔ245 (an overproducer of $\sigma^{70(aa130-374)del}$, referred to henceforth as $\sigma^{\Delta 245}$ [Kumar *et al.*, 1995b] (Fig 3.1), in which the *Xba* I–*Not* I–*Nsi* I fragment is replaced with a *Sac* I restriction site (by using oligonucleotides RH7 and RH8). pTK310, a derivative of pGEMD (an overproducer of σ^{70}, [Igarashi and Ishihama, 1991]), was constructed in exactly the same way.

To facilitate the cloning of inserts carrying mutations, the *Bam*H I–*Xho* I regions of pHW111 (Fig. 2.1) [R. Hayward and H. S. Williamson, unpublished], pTK210 and pTK310 were replaced with a *Bam*H I–*Not* I–*Xho* I linker (oligonucleotides N4356 and

N4357). The resulting plasmids were named pTK111, pTK211 and pTK311 (Fig. 3.1).

Site-directed mutagenesis creating single and double amino acid residue substitutions in σ^{70} (Y421C, Y425C, Y425C W433A, S428C, K502C, and S517C) was carried out by the Kunkel method (as described in 2.4.14), with bacteriophage M13mp18-14 (M13mp18 with its *Sac* I–*Pst* I segment replaced with the *Sac* I–*rpoD*–*Hgi*A I fragment of pSAK3 [Kumar *et al.*, 1993]) as a source of *rpoD*$^+$ single-stranded template DNA. Other mutations (Q437C, S442C, and T440C) were introduced by a PCR method (2.4.14), using pSAK15–70/32 [Kumar *et al.*, 1995a] as a source of *rpoD* DNA. The second primer for PCR mutagenesis was RH6, which contains the *Xho* I site of *rpoD*. The mutagenic primers are listed in Table 3.1.

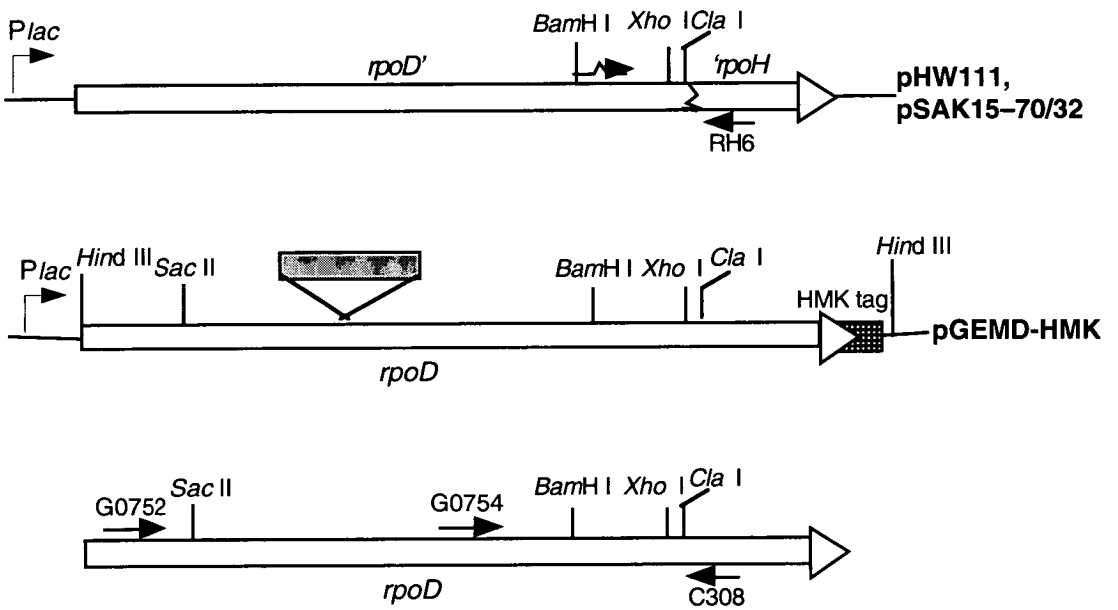


Figure 3.1. Strategy for site directed mutagenesis by PCR.
⟿ – mutagenising primer. Restriction sites used for cloning mutagenised inserts are also shown. Primer used for sequencing are shown in the bottom figure.

The DNA sequences of the *Bam*H I–*Xho* I fragments from the plasmids and M13 phages produced by either methods were verified by sequencing using primers C308, G0752 and G0754 (Fig. 3.1).

Their *Bam*H I–*Xho* I or *Bam*H I–*Cla* I fragments were cloned into pTK111 and pTK211, and some of them into pTK311, to produce the pTK118–pTK124, pTK218–pTK224, and pTK318–pTK324 series of plasmids (Table 3.1). The presence of the inserts was confirmed by restriction analysis using *Not* I and also *Nde* I, and, in the cases of pTK213, pTK215, pTK217, pTK218 and pTK219, by sequencing the inserted region using primer G0754.

Table 3.1. *rpoD* mutations and plasmids.

Mutation	Region of the mutation	Primer used	Nucleotide substitution	$\sigma^{70/32}$ hybrid plasmid	$\sigma^{\Delta 245}$ overpro-ducing plasmid	σ^{70} overpro-ducing plasmid	$\sigma^{\Delta 245}$-HMK overpro-ducing plasmid
wt				pHW111	pTK210	pTK310	
Y421C	2.3	NS14	TAC→TGC	pTK124	pTK224	pTK324	pTK424
Y425C	2.3	NS12	TAC→TGC	pTK122	pTK222	pTK322	pTK422
Y425C W433A	2.3	NS13	TAC→TGC TGG→GCG	pTK123	pTK223	pTK323	pTK423
S428C	2.3	NS10	TCC→TGC	pTK120	pTK220	pTK320	pTK420
Q437C	2.4	RH1	CAG→TGC	pTK113	pTK213		pTK413
T440C	2.4	RH2	ACC→TGC	pTK117	pTK217		pTK417
S442C	2.4	RH3	TCT→TGT	pTK115	pTK215		pTK415
K502C	3.1	RH4	AAA→TGC	pTK118	pTK218		pTK418
S517C	3.2	RH5	TCG→TGC	pTK119	pTK219		pTK419

Sac II–*Xho* I fragments of pTK213, pTK215, pTK217, pTK218, pTK219, pTK220, pTK222, pTK223, and pTK224 were cloned into pGEMD-HMK. pGEMD-HMK is pGEMD with a ~2 kb insert between the *Sac* II and *Bam*H I sites of *rpoD*, and an HMK (human muscle kinase)-tag coding sequence attached to the downstream terminus of the coding sequence of *rpoD*, and was kindly provided by H. Nagai. This resulted in the pTK413–pTK424 series of plasmids, which are $\sigma^{\Delta 245}$-overproducers with the corresponding mutations

and with a sequence encoding SARRASVA (RRASV is the amino acid sequence of the HMK-tag) attached to the downstream terminus of the gene:

*Eco*O47 III *Hind* III
*rpoD agcgct*AGAAGGGCAAGTGTTgcat*aaGCTT*
S A **R R A S V** A *

3.2. The *in vivo* activities of sigma subunits with single cysteine replacements.

The effects of the mutations on the activity of σ were analysed *in vivo* using the hybrid sigma assay [Kumar *et al.*, 1995a] described in 2.5.10. The results of the assays in which pTK113–124 were tested against the pHW111 control are presented in Fig. 3.2.

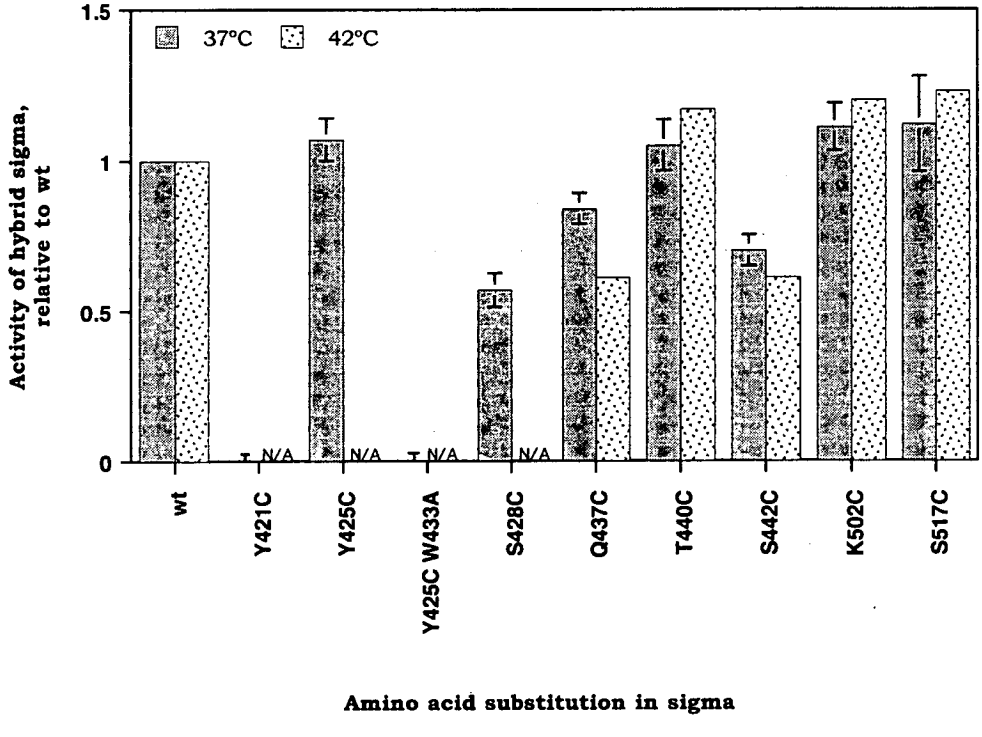


Figure 3.2. Activities of hybrid $\sigma^{70/32}$ subunits. The bars indicate ratios of activities of the hybrid sigmas with the designed cysteine replacements in the σ^{70} portion, to the activity of the hybrid sigma $\sigma^{70/32}$ with the wild-type sequence of σ^{70} and σ^{32}. Left bars are activities at 37°C, right ones are at 42°C. The activities of $\sigma^{70/32\ Y421C}$ and $\sigma^{70/32\ Y425C\ W433A}$ at 37°C are zero. Error bars show the standard error of the ratios for measurements at 37°C. The data for 42°C are the mean of two measurements. *N/A* – these proteins were not assayed at 42°C.

As seen from Fig. 3.2, the hybrid $\sigma^{70/32}$ subunits carrying single cysteine replacements Y425C, T440C, K502C, and S517C exhibit activities 1.05–1.12 that of the hybrid $\sigma^{70/32}$ with the wild-type sequence at the hybrid promoter at 37°C. The hybrid $\sigma^{70/32}$ subunits with the S428C, S442C, and Q437C substitutions show activities lower than that of the wild type $\sigma^{70/32}$ with values of 0.57–0.84 that of $\sigma^{70/32}$. However, the hybrid $\sigma^{70/32}$ with Y421C and double Y425C W433A substitutions are inactive at 37°C. This was quite unexpected for Y421C hybrid sigma, because $\sigma^{70(R588H\ Y421C)}$ retained 45% activity compared to $\sigma^{70(R588H)}$ at mutated promoter *ant↓–33A in vivo* [Waldburger and Susskind, 1994]. On the other hand, despite of examples of functionality of sigmas with W433 substitutions, Juang *et al.* [1995] showed that $\sigma^{A(W192A)}$ of *B. subtilis* was defective in nucleation and propagation of the melted region of *PtrnS* promoter *in vitro* (W192 of σ^A corresponds to W433 of σ^{70}), and overexpression of $\sigma^{A(W192A)}$ *in vivo* was lethal to cells [Rong and Helmann, 1994], as was mentioned in Chapter 1. It is also possible that, even if Y425C and W433A substitutions separately would not make σ^{70} inactive, the double substitution could severely damage its ability of promoter melting as these two residues are present on the same side of α-helix in the crystal structure of a portion of σ^{70} [Malhotra *et al.*, 1996].

3.3. Hydrophobicity of $\sigma^{\Delta 245\text{-}HMK}$ and its derivatives with single cysteine replacements.

All the $\sigma^{\Delta 245}$ overproduction plasmids (pTK218–pTK224 series) were checked for their ability to overproduce proteins of the expected size (2.5.3). As seen from the Fig. 3.3, all the plasmids express proteins ($\sigma^{\Delta 245}$ and its derivatives) of the same apparent molecular mass of 59–62 kDa, whereas their calculated molecular mass is 42 kDa (the estimate of $\sigma^{\Delta 245}$ mobility of Kumar *et al.* [1995b] is 55 kDa). The anomalous mobility of $\sigma^{\Delta 245}$ on SDS–PAGE is consistent with anomalous SDS–PAGE mobility of σ^{70} (86–95 kDa) and σ^A (55 kDa) [Shorenstein and Losick, 1973], with their calculated molecular masses being 70 and 43 kDa, respectively.

Four $\sigma^{\Delta 245}$ proteins with single cysteine replacements, namely with $\sigma^{\Delta 245\text{-HMK Y425C}}$, $\sigma^{\Delta 245\text{-HMK S428C}}$, $\sigma^{\Delta 245\text{-HMK Q437C}}$, and $\sigma^{\Delta 245\text{-HMK K502C}}$, were overproduced and purified according to the protocol described in 2.5.3.–7. $\sigma^{70\text{-HMK}}$ [Nagai and Shimamoto, 1997] and $\sigma^{\Delta 245\text{-HMK}}$ were purified by H. Nagai.

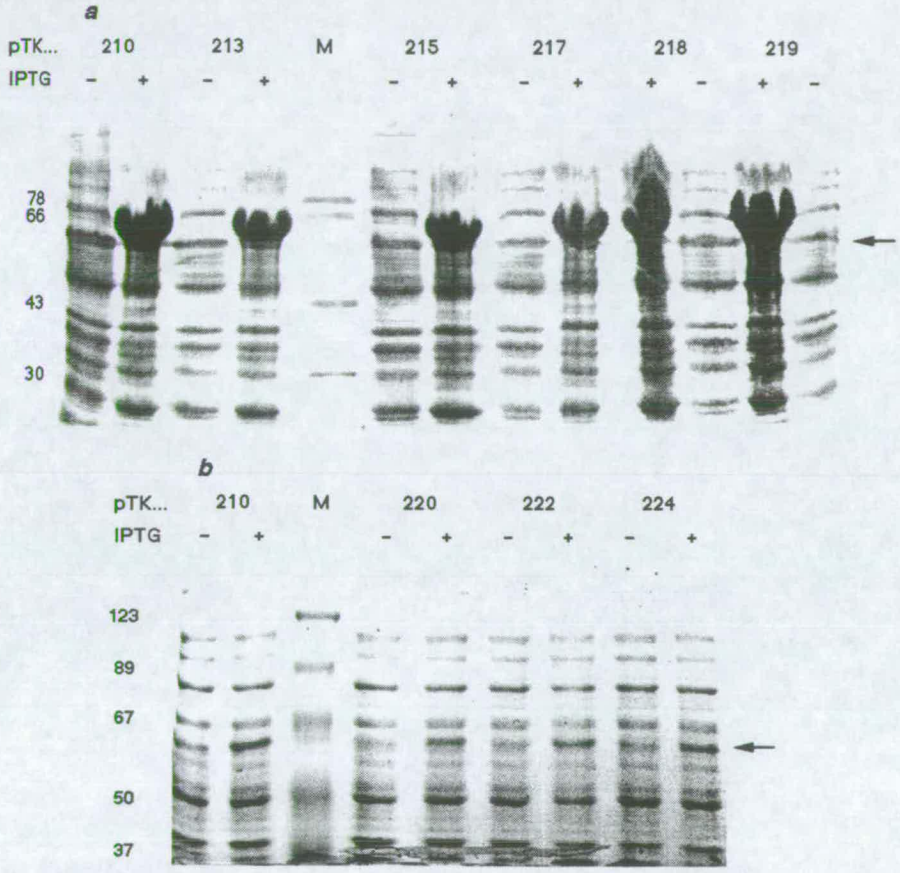


Figure 3.3. 10% SDS–PAGE of BL23(λDE3)·plysS total cell extracts carrying pTK210 or its derivatives pTK213, pTK215, pTK217, pTK218. pTK219, pTK220, pTK222, and pTK224. The cells were grown to A^{600} 0.6, were not induced (IPTG –) or induced (IPTG +) with 125 mg l^{-1} (gel *a*) or 25 mg l^{-1} (gel *b*) IPTG and incubated for further 3 (gel *a*) or 1 (gel *b*) hour, before harvesting. Arrows indicate bands of 59–62 kDa corresponding to the induced proteins, derivatives of $\sigma^{\Delta 245}$. *M* – molecular mass markers (kDa), their corresponding molecular masses are on the left of the gels.

Overproduced $\sigma^{\Delta 245\text{-}HMK}$ and its single cysteine replacement derivatives form inclusion bodies in *E. coli* cells. These inclusion bodies were purified and denatured in 6 M guanidine chloride solution. During the refolding step, when the guanidine chloride solutions were dialysed against TGEB buffer and further against 0.5 M ammonium sulphate in TGEB, the solutions became turbid even during the early steps of dialysis indicating of hydrophobicity of $\sigma^{\Delta 245\text{-}HMK}$ and its derivatives. A novel procedure [Shimamoto *et al.*, 1998] was applied at this stage, in which a protein solution containing 6 M guanidine chloride was diluted 2-fold with glycerol to prevent precipitation during the dialysis, and dialysed against 10 volumes of 1:3 (v/v) mixture of TEB and glycerol at 4°C during 10–16 hours, with constant rotating and 3–4 changes of the outer solution during the dialysis. The dialysis was monitored by the conductivity of the outer dialysis buffer. The dialysed solution was transparent, whereas when the dialysis was performed against mixture of 1 part of TEB buffer and 1 part of glycerol, the dialysed solution became turbid, indicative of protein precipitation.

After the dialysis and centrifugation the protein solution was rapidly diluted against 10 volumes of 0.3 M ammonium sulphate in TGEB buffer, and the proteins were precipitated at 4°C by adding ammonium sulphate to 1 M. The proteins were dissolved in 0.2 M NaCl in TGEB buffer. These solutions, with estimated concentrations of 10–14 mg ml^{-1}, become opaque and then gelled after 3–4 days at 4°C. Addition of Tween-20 to 0.025% reverses the polymerisation process.

The protein solutions were subjected to fractionation on a Phenyl Toyopearl ($\sigma^{\Delta 245\text{-}HMK}$) (H. Nagai) or a POROS-HQ/M ($\sigma^{\Delta 245\text{-}HMK\ Y425C}$ and $\sigma^{\Delta 245\text{-}HMK\ S428C}$) chromatographic columns. $\sigma^{\Delta 245\text{-}HMK}$ could not been eluted from the hydrophobic Phenyl Toyopearl column by the 1.5 to 0 M ammonium sulphate gradient in TGEB buffer (H. Nagai). The fractionation of $\sigma^{\Delta 245\text{-}HMK\ Y425C}$ and $\sigma^{\Delta 245\text{-}HMK\ S428C}$ on POROS-HQ/M column with 0.2 to 1 M NaCl gradient in TGEB buffer with 0.025% Tween-20 was poor, these

proteins were eluted together with the whole mixture of injected proteins (Fig. 3.4).

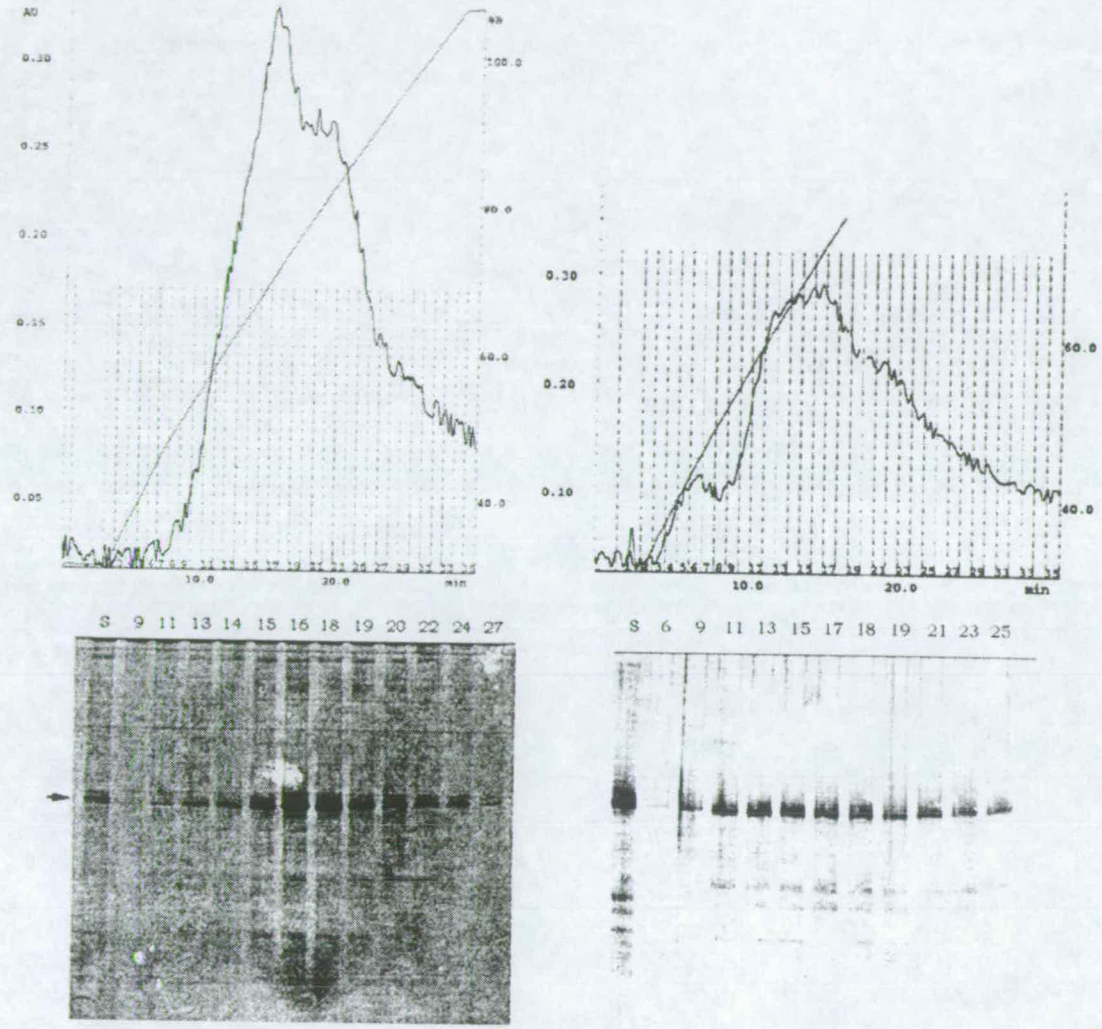


Figure 3.4. Chromatograms of $\sigma^{\Delta 245\text{-HMK S428C}}$ and $\sigma^{\Delta 245\text{-HMK Y425C}}$ and SDS–PAGE of their chosen fractions underneath. POROS-HQ/M column, loaded at flow rate 2 ml min^{-1}, elution by 0.2 to 1 M NaCl gradient in TGEB buffer with 0.025% Tween-20 at flow rate 0.5 ml min^{-1}. 0.5 µl of the shown fractions are separated on 10% SDS–PAGE and the gels are reverse-stained. Several protein bands present in the loading solution (*S*) co-fractionate with $\sigma^{\Delta 245\text{-HMK S428C}}$ and $\sigma^{\Delta 245\text{-HMK Y425}}$.

Considering the poor fractionation of re-folded $\sigma^{\Delta 245\text{-HMK}}$ and its derivatives on the anion-exchange column and poor solubility in TGEB buffer due to high hydrophobicity, an attempt was made to

use another technique of fractionating $\sigma^{\Delta 245\text{-HMK}}$ derivatives on a hydrophobic Phenyl-Toyopearl (medium) column in a denatured form (chapter 2.5.6).

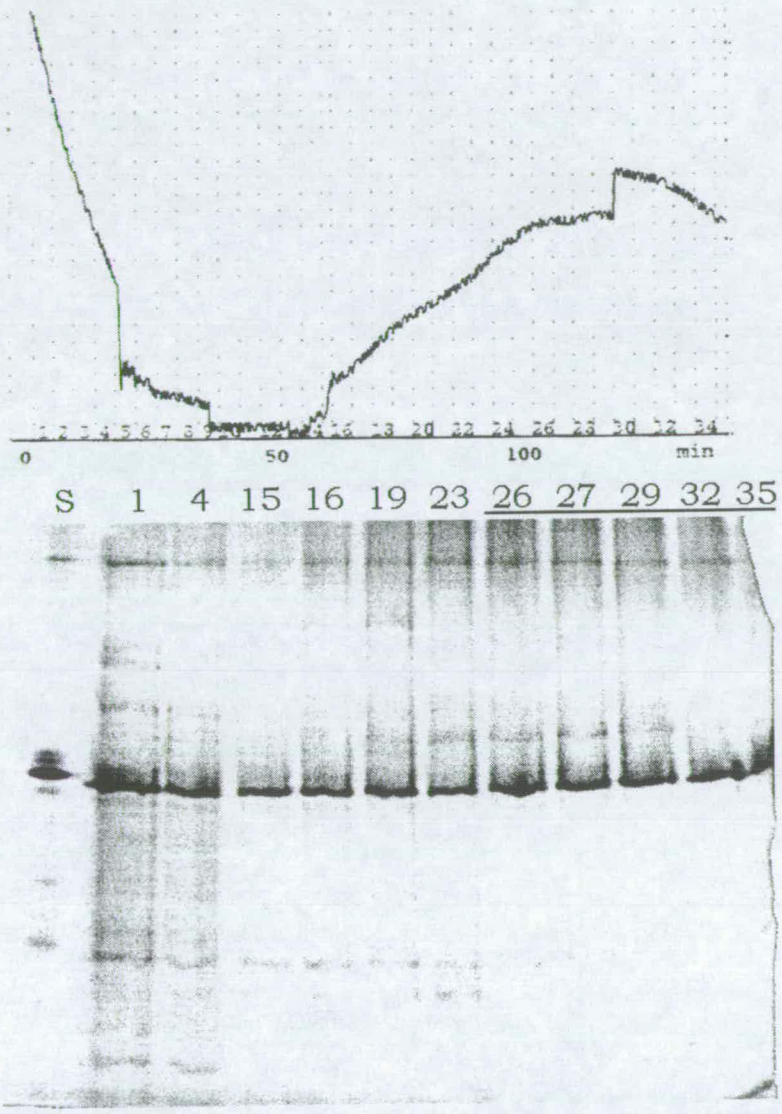


Figure 3.5. The chromatogram of $\sigma^{\Delta 245\text{-HMK Q437C}}$ mixture and SDS–PAGE of chosen fractions. Phenyl Toyopearl (Tosoh, Tokyo) column was equilibrated with 0.9 M ammonium sulphate 6 M guanidine chloride in TGEB buffer, elution by 0.9 to 0 M ammonium sulphate gradient in 6 M guanidine chloride/TGEB buffer. 0.5 µl of several fractions are separated on 10% SDS–PAGE and the gel was reverse-stained. *S* – loading solution; fractions 26–35 (underlined) were pooled.

Protein solution ($\sigma^{\Delta 245\text{-HMK Y425C}}$, $\sigma^{\Delta 245\text{-HMK S428C}}$, $\sigma^{\Delta 245\text{-HMK Q437C}}$, or $\sigma^{\Delta 245\text{-HMK K502C}}$) in 6 M guanidine chloride/TGEB buffer were loaded on the column and eluted with a

0.9 to 0 M gradient of ammonium sulphate in 6 M guanidine chloride/TGEB buffer (Fig. 3.5).

Fractions containing the least impurities were pooled, diluted 2-fold with glycerol to prevent precipitation during the dialysis, and dialysed as before against 10 volumes of 1:3 (v/v) mixture of TEB and glycerol. The protein concentrations in the resulting solutions was determined by BCA protein assay (2.5.7). The proteins were stored at -20°C and rapidly diluted in reaction buffer before usage.

The dialysis against 75% glycerol is a novel technique for solubilisation of hydrophobic proteins (see Appendix). However, at such high glycerol concentration some the protein folding may not be complete. Measurements of activity of $\sigma^{\Delta 245\text{-HMK}}$ and its derivatives described below show that at least some fraction of the proteins is active *in vitro*.

3.4. Reversibility of oligomerisation of $\sigma^{\Delta 245\text{-HMK}}$ and of its single cysteine replacement derivatives.

In order to determine the size of the purified $\sigma^{\Delta 245\text{-HMK}}$ it was subjected to gel filtration on Superose 6 column. The chromatogram is shown on Fig. 3.6, *a*. Each eluted fraction was analysed by SDS–PAGE (Fig. 3.6, *b*). The major peak of eluted $\sigma^{\Delta 245\text{-HMK}}$ corresponds to the void volume (fraction 11) of this column, which indicates that the number of molecules in aggregates in this fraction is not less than 45 (the column was calibrated by H. Nagai), whereas $\sigma^{70\text{-HMK}}$ and *B. subtilis* σ^{A} with a His-tag on its *C*-terminus are eluted at volumes corresponding to 117 and 62 kDa, (the data and column calibration by H. Nagai), which is molecular weight more than their monomer (70 and 43 kDa), but less than their dimer forms.

The right shoulder of the major peak on the chromatogram in Fig. 3.6, *a* (fractions 12–20) indicates the presence of oligomers of sizes up to about 650 kDa, that is not less than 12-mers. The peak at fractions 27 corresponds to molecular sizes about 53 kDa, and peak

at fraction 35 is not a protein peak (presumably β-mercaptoethanol is eluted and detected at 280 nm in fraction 35).

Fraction 11 (eluted with the void volume of the column) was kept at 4°C for 2 hours and then re-injected onto the Superose 6 column (Fig. 3.6, *b*). The eluted fractions were analysed by SDS–PAGE, and the relative intensity of the $\sigma^{\Delta 245\text{-HMK}}$ was measured by densitometry (Fig. 3.6, *c* and *d*). This was done to show which peaks on the chromatogram actually correspond to protein peaks. Fractions 22 (molecular mass about 400 kDa) and 29 (approximately 29 kDa, presumably the monomer form of $\sigma^{\Delta 245\text{-HMK}}$) contain $\sigma^{\Delta 245\text{-HMK}}$ too, and this indicates that during 2 hours at 4°C in the elution buffer (0.2 M NaCl, 0.025% Tween-20 in TGEB) the equilibrium between high size oligomer and oligomers of smaller sizes (8–10-mer) and monomer, is shifted towards the monomer form, i. e. the oligomerisation of $\sigma^{\Delta 245\text{-HMK}}$ is reversible. Similar chromatograms were obtained for $\sigma^{\Delta 245\text{-HMK Y425C}}$ and $\sigma^{\Delta 245\text{-HMK S428C}}$.

When the core RNA polymerase ($\alpha_2\beta\beta'$) is subjected to chromatography on the Superose 6 column under the same conditions, it is eluted in fraction 22 which corresponds to approximate molecular mass of 450 kDa (calculated 378 kDa). The mixture of preincubated (10 minutes at 37°C) core enzyme and $\sigma^{\Delta 245\text{-HMK}}$ (1:4) is eluted in fraction 20, which corresponds to approximate molecular mass 500–550 kDa. This indicates binding of $\sigma^{\Delta 245\text{-HMK}}$ to the core enzyme, however, the number of $\sigma^{\Delta 245\text{-HMK}}$ molecules bound is unclear from this experiment. The peak of oligomeric form of $\sigma^{\Delta 245\text{-HMK}}$ is eluted in the void volume (fraction 11).

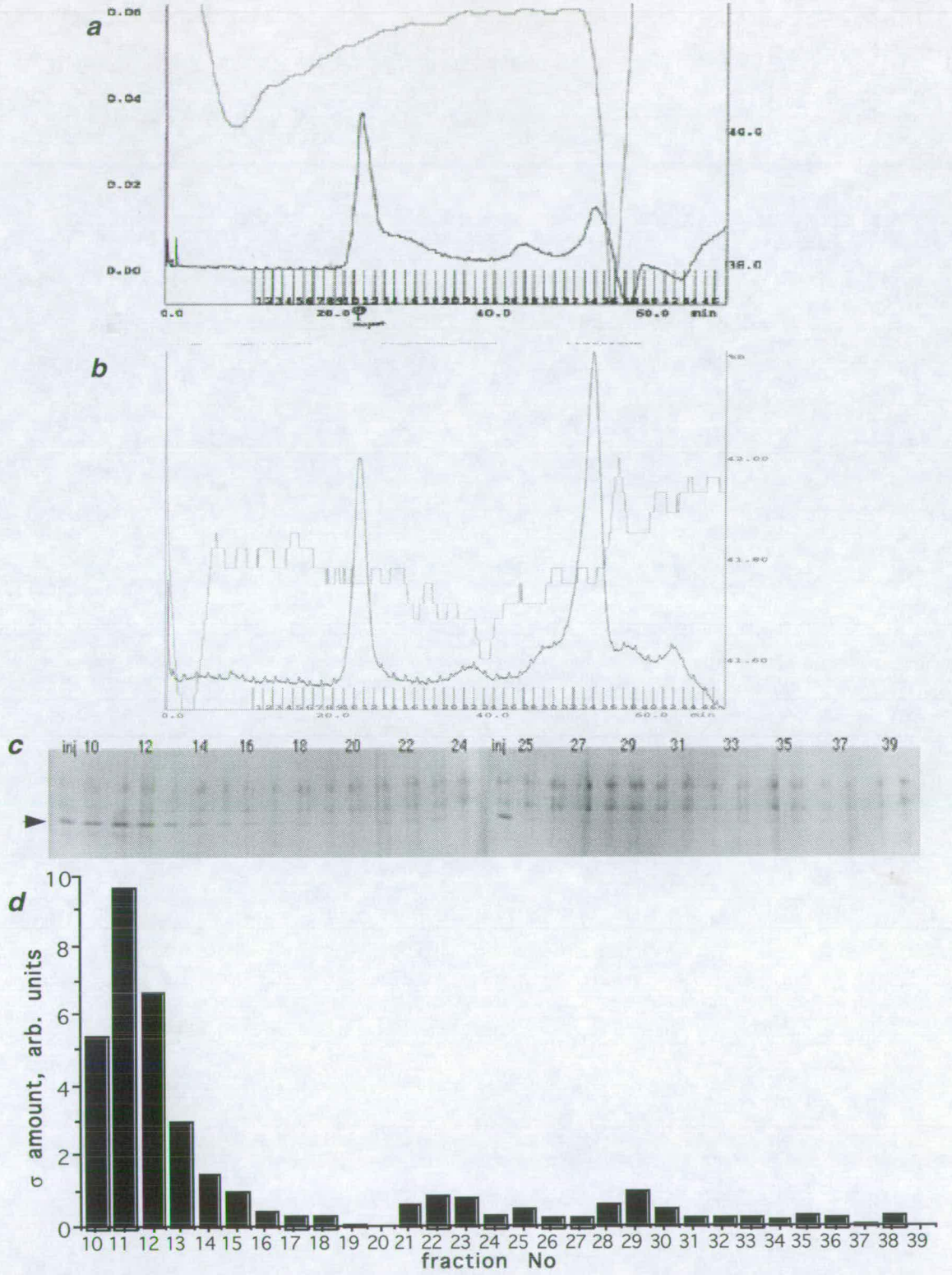


Fig. 3.6. *a*) Chromatogram of 30 μg $\sigma^{\Delta 245\text{–HMK}}$ in 20 μl of elution buffer 0.2 M NaCl, 0.025% Tween 20 in TGEB, 40 μl min^{-1}, 10°C, total back pressure 1.2 to 1.4 MPa. Detection at 280 nm (bottom line) and relative conductivity (top line) are shown. *b)* Chromatogram of fraction 11 of the previous chromatography, under the same conditions. The fraction was kept at 4°C for 2 hours before re-applying. *c)* SDS–PAGE of fractions 10–39 of the chromatography *b*; *inj*, 3 μl of the injected fraction 11. The gel was reverse-stained, the relative amount of $\sigma\Delta^{245\text{–HMK}}$ in every lane was determined by densitometry and plotted on the graph *d*.

3.5. Activity of E·$\sigma^{\Delta 245\text{-HMK}}$ and the derivatives *in vitro*.

$\sigma^{\Delta 245\text{-HMK}}$ and its four single-cysteine replacement derivatives were titrated against the core enzyme to determine the saturating molar ratio of the core enzyme to σ subunits at which transcription occurs. This assay measured transcription *in vitro* using the T7A1 promoter template. The analysis was performed with 30 nM promoter template, 40 nM core enzyme and σ subunits concentrations from 40 to 800 nM. The working ratio of the core enzyme to a σ subunit at which saturation of full-length transcripts occurs is 1:4 for $\sigma^{\Delta 245\text{-HMK}}$ and $\sigma^{\Delta 245\text{-HMK Q437C}}$, 1:8 for $\sigma^{\Delta 245\text{-HMK K502C}}$, 1:10 for $\sigma^{\Delta 245\text{-HMK S428C}}$, and 1:15 for $\sigma^{\Delta 245\text{-HMK Y425C}}$. In these proportions the σ subunits were further used in the reconstituted holoenzymes; the molar ratio of the core to σ^{70} and $\sigma^{70\text{–HMK}}$ was 1:4.

3.6. T7A1 DNA promoter binding and open complex formation.

To determine how $\sigma^{\Delta 245\text{-HMK}}$ and its four single-cysteine replacement derivatives bind T7A1 promoter DNA and form open complexes, gel shift assays of the reconstituted holoenzymes (from 0 to 100 nM) with 10 nM of T7A1 promoter DNA was performed. Heparin was added to the reaction mixture before the loading to bind all non-specifically bound to the DNA holoenzyme, as well as the specific promoter–polymerase complexes non-resistant to heparin (closed complexes). The dried gels were subjected to phosphoimaging analysis, and the amount of bound DNA was plotted against the holoenzyme concentration (Fig. 3.7 and 3.8). It can be seen from the figure that E·$\sigma^{\Delta 245\text{-HMK Y425C}}$ and E·$\sigma^{\Delta 245\text{-HMK S428C}}$, which have σ subunits with cysteine replacements in region 2.3, have reduced ability to form heparin-resistant open complex, whereas E·$\sigma^{\Delta 245\text{-HMK}}$, E·$\sigma^{\Delta 245\text{-HMK Q437C}}$, and E·$\sigma^{\Delta 245\text{-HMK K502C}}$ all form heparin-resistant open complex with the core enzyme similar to E·σ^{70}. The apparent K_D for E·$\sigma^{\Delta 245\text{-HMK Y425C}}$, which is hard to determine here, is the lowest to compare to all the other enzymes tested here.

It has also been shown that the molar ratio of a σ subunit to the core in case of $\sigma^{\Delta 245\text{-HMK}}$ and its four single-cysteine replacement derivatives is 1:1. Gel-shift assays of the holoenzymes and of $E{\cdot}\sigma^{70}$ with T7A1 promoter DNA with and without the addition of heparin were performed. The gel bands were visualised using a BAS2000 Image Analyser, and the shifted DNA bands were cut out of the native gel and subjected to SDS–PAGE, together with the known amounts of $E{\cdot}\sigma^{70}$ holoenzyme. The gel was reverse-stained, and the ratio of intensity of σ to α subunit bands in each track was measured. The data indicated that the amount of σ subunit per core enzyme in the heparin-resistant complex of the holoenzyme with T7A1 DNA was close to 1 for all the holoenzymes studied.

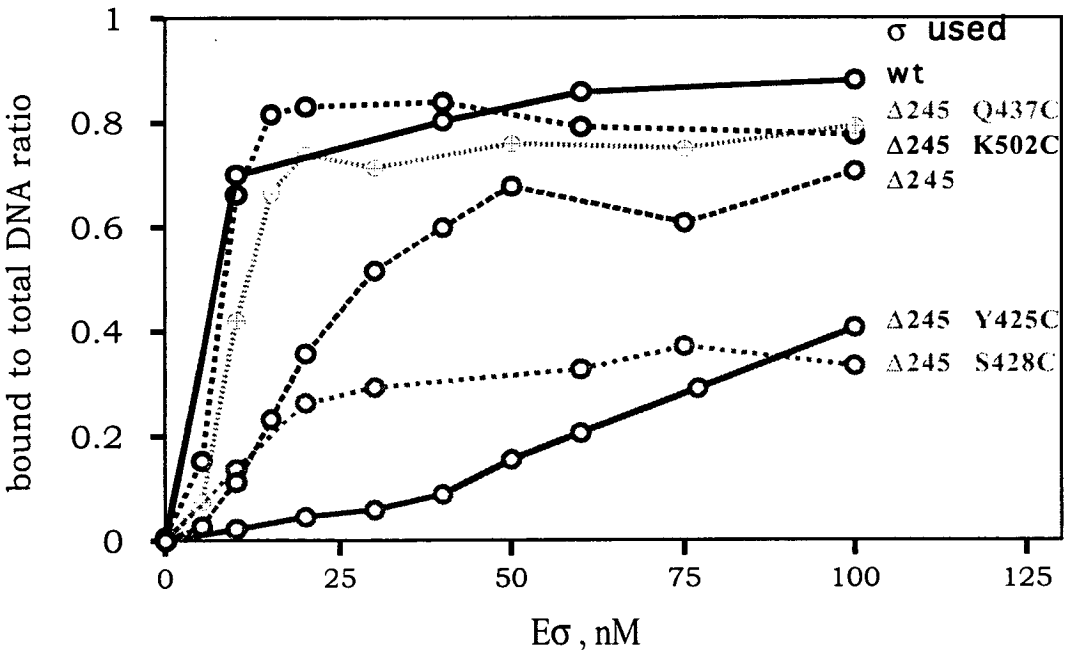
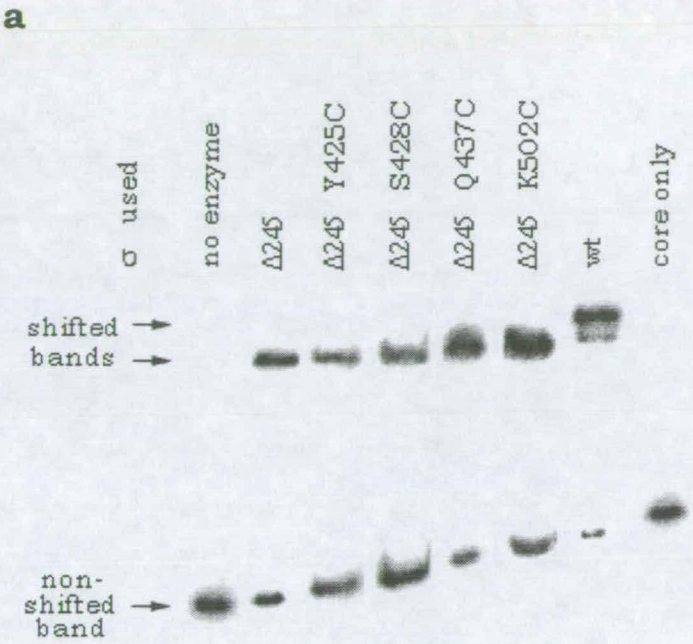


Figure 3.7. Data of gel-shift assays of heparin-resistant complexes of holoenzymes (0 to 100 nM) to 10 nM T7A1 promoter.



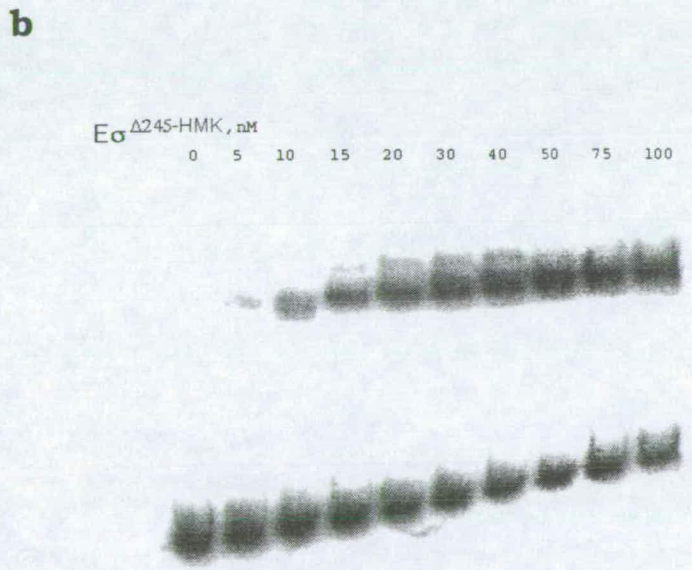


Figure 3.8. *a)* The gel-shift assays for holoenzyme concentrations 100 nM in Fig. 3.7. *b)* The gel-shift assay corresponding to $E{\cdot}\sigma^{\Delta 245\text{-HMK}}$ data in Fig. 3.7.

3.7. The activity of E·$\sigma^{\Delta 245\text{-HMK}}$ and the derivatives from T7A1 and λP_R promoters *in vitro*.

The activity of the reconstituted holoenzymes was studied on two promoters – T7A1 and λP_R. The results for the T7A1 promoter where the DNA is present in excess are shown in Fig. 3.9. In the presence or absence of heparin, the HMK tag on the *C*-terminus of σ^{70} does not affect the transcription from the T7A1 promoter. However, the transcription by E·$\sigma^{\Delta 245\text{-HMK}}$ is only about 0.18 of that by E·σ^{70}, in presence or absence of heparin. Therefore, on the T7A1 promoter E·$\sigma^{\Delta 245\text{-HMK}}$ is less efficient than E·σ^{70} and E·$\sigma^{70\text{-HMK}}$ in the steps following promoter opening. Transcription from T7A1 promoter by holoenzymes with single cysteine replacements of $\sigma^{\Delta 245\text{-HMK}}$ does not differ much for the K502C substitute, is somewhat lower for the Y425C and S428C substitutes, but is significantly lower for $\sigma^{\Delta 245\text{-HMK Q437C}}$. The latter result is, although consistent with the earlier observations that Q437 is involved in promoter recognition at position –12 (described in Chapter 1), is inconsistent with the relatively high rate of formation of a heparin-resistant complex with this promoter described in the previous section. If Q437 is involved in binding single-stranded non-template strand with specificity to a nucleotide at position –12 of a promoter (T in a consensus promoter, G in case of T7A1 and λP_R promoters), then the reduction of transcription level is expected to be due to the defect in binding the non-template strand in the open complex, which is presumably observed as a heparin-resistant complex in gel shift assays. In our case it seems that the heparin-resistant (open) complexes E·$\sigma^{\Delta 245\text{-HMK Q437C}}$ forms with T7A1 promoter are different from a productive open complex and are defective in both productive and abortive transcription.

The ratio of abortive (6-mer) to the full-length transcript under this conditions (in the presence of heparin) is similar for E·σ^{70}, E·$\sigma^{70\text{-HMK}}$, E·$\sigma^{\Delta 245\text{-HMK}}$ and its derivatives (Fig. 3.9*c*). In a multiple-round transcription, when heparin is not added to the reaction, this ratio is similar again for E·σ^{70}, E·$\sigma^{70\text{-HMK}}$ and E·$\sigma^{\Delta 245\text{-HMK}}$.

The transcription from λP_R promoter was conducted under two conditions – with the excess of RNA polymerase and with the equimolar amounts of DNA and enzyme. The results are presented in Fig. 3.10. When the enzyme is in excess (10 nM promoter DNA and 100 nM of RNA polymerase) the deletion of 245 amino acid residues from σ^{70} does not make any difference for the level of transcription from λP_R promoter. Only $E{\cdot}\sigma^{\Delta 245\text{-HMK Q437C}}$ exhibits the low level of transcription of all the holoenzymes tested, about a half that of $E{\cdot}\sigma^{70}$.

The ratio of abortive to the full-length transcript is lower for $E{\cdot}\sigma^{\Delta 245\text{-HMK}}$ and its derivatives than that for $E{\cdot}\sigma^{70\text{-HMK}}$ under the conditions of excess of a holoenzyme. This can indicate that the abortive complex $E{\cdot}\sigma^{\Delta 245\text{-HMK}}_{abort}{\cdot}P$ is weaker than the productive complex $E{\cdot}\sigma^{\Delta 245\text{-HMK}}_{prod}{\cdot}P$ under these conditions, while σ^{70} forms more abortive complexes per productive complex.

Under conditions when holoenzymes are not in excess, $E{\cdot}\sigma^{\Delta 245\text{-HMK}}$ and its derivatives $E{\cdot}\sigma^{\Delta 245\text{-HMK S428C}}$ and $E{\cdot}\sigma^{\Delta 245\text{-HMK K502C}}$ do not differ significantly in their ability to transcribe from this promoter, comparative to $E{\cdot}\sigma^{70}$ and $E{\cdot}\sigma^{70\text{-HMK}}$. $E{\cdot}\sigma^{\Delta 245\text{-HMK Y425C}}$ and $E{\cdot}\sigma^{\Delta 245\text{-HMK Q437C}}$ show levels of the full-length transcripts much less than $E{\cdot}\sigma^{\Delta 245\text{-HMK}}$. The data for single cysteine substitutes of $\sigma^{\Delta 245\text{-HMK}}$ are from the single experiment, and need confirmation. The level of transcription by $E{\cdot}\sigma^{\Delta 245\text{-HMK Q437C}}$ is, however, consistent with the data where the enzyme was in excess, and with the data from transcription from T7A1 promoter described above.

3.8. Abortive synthesis from λP_R promoter.

As it has been shown previously, $E{\cdot}\sigma^{70}$ is able to produce abortive transcripts for some time *in vitro* after the productive transcription had been completed [Kubori and Shimamoto, 1996]. An example of this is demonstrated on the first gel of Fig. 3.11. On λP_R promoter, the production of the full length 32 nt transcript ceases after 5 minutes of the reaction, while the production of abortive transcripts continues beyond 20 minutes of reaction. The same is

true for $E{\cdot}\sigma^{70\text{-HMK}}$. However, for $E{\cdot}\sigma^{\Delta 245\text{-HMK}}$, although the abortive complexes survive a little longer than the productive ones, the synthesis of both abortive and productive transcripts stops by the tenth minute of the reaction (Fig. 3.11). This means that the abortive complex $E{\cdot}\sigma^{\Delta 245\text{-HMK}}{}_{abort}{\cdot}P$ does not stay on the promoter for as long as $E{\cdot}\sigma^{70}{}_{abort}{\cdot}P$ or $E{\cdot}\sigma^{70\text{-HMK}}{}_{abort}{\cdot}P$, and the reaction of formation of an abortive complex is reversed back to $E{\cdot}\sigma + P$, whereupon holoenzyme $E{\cdot}\sigma$ is trapped by heparin.

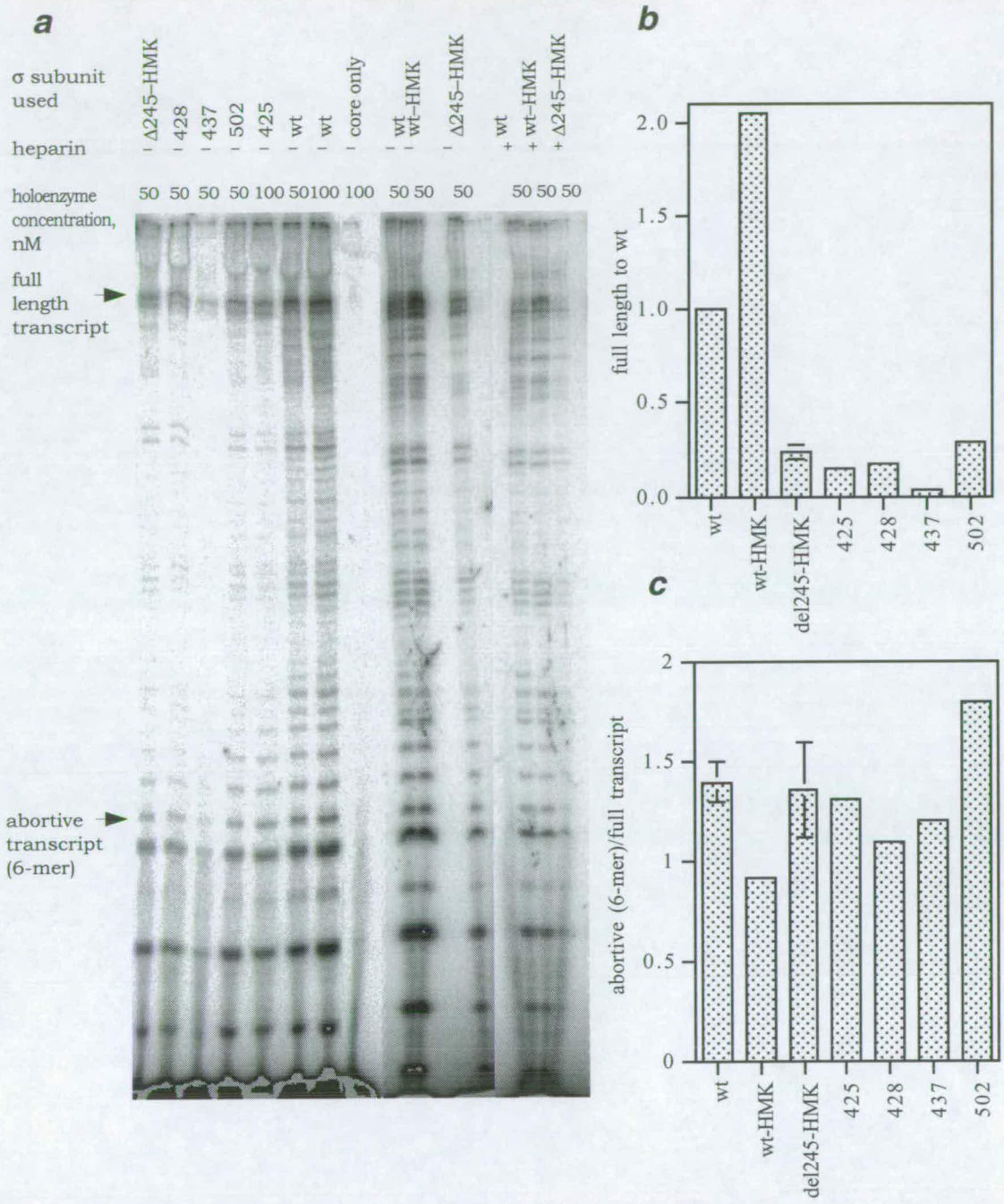


Figure 3.9. a) and Transcription reactions from T7A1 promoter DNA. DNA concentration 200 nM (tracks 1–7) and 100 nM (tracks 8–13). The σ subunits used in each reconstituted holoenzyme, presence or absence of heparin and concentrations of holoenzymes are shown on the top. 425, 428, 437, and 502 stand for $\sigma^{\Delta 245\ HMK}$ with corresponding amino acid residue substituted for cysteine. The ratio of abortive transcript (6-mer) to the full length transcript is close to that for the wild type enzyme for all the enzymes used here. b) The amount of full length transcript produced relative to the wild type holoenzyme in the absence of heparin. c) The amount of abortive (6-mer) transcript relative to the full length transcript (in the absence of heparin).

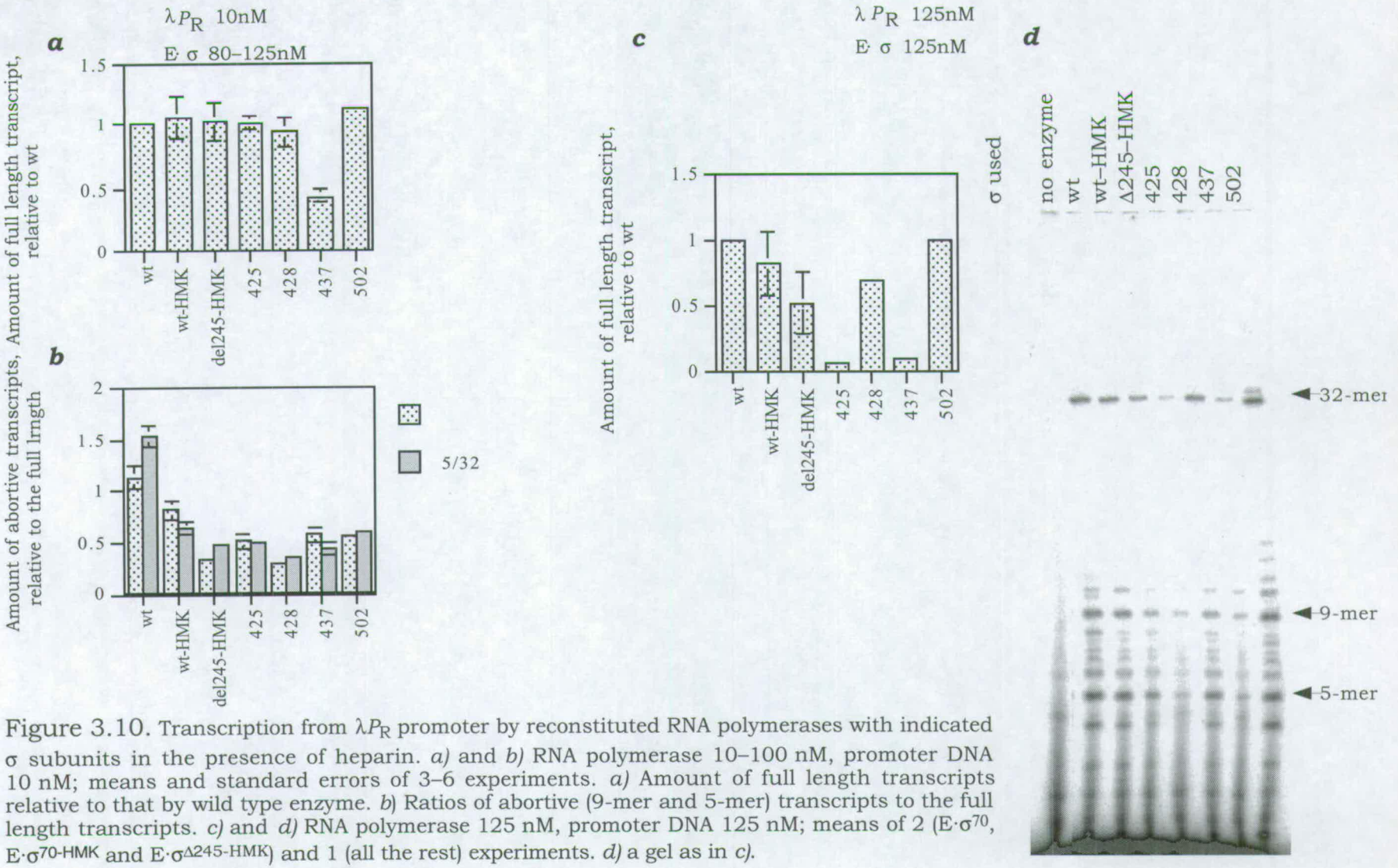


Figure 3.10. Transcription from λP_R promoter by reconstituted RNA polymerases with indicated σ subunits in the presence of heparin. *a)* and *b)* RNA polymerase 10–100 nM, promoter DNA 10 nM; means and standard errors of 3–6 experiments. *a)* Amount of full length transcripts relative to that by wild type enzyme. *b*) Ratios of abortive (9-mer and 5-mer) transcripts to the full length transcripts. *c)* and *d)* RNA polymerase 125 nM, promoter DNA 125 nM; means of 2 ($E\cdot\sigma^{70}$, $E\cdot\sigma^{70\text{-HMK}}$ and $E\cdot\sigma^{\Delta 245\text{-HMK}}$) and 1 (all the rest) experiments. *d)* a gel as in *c)*.

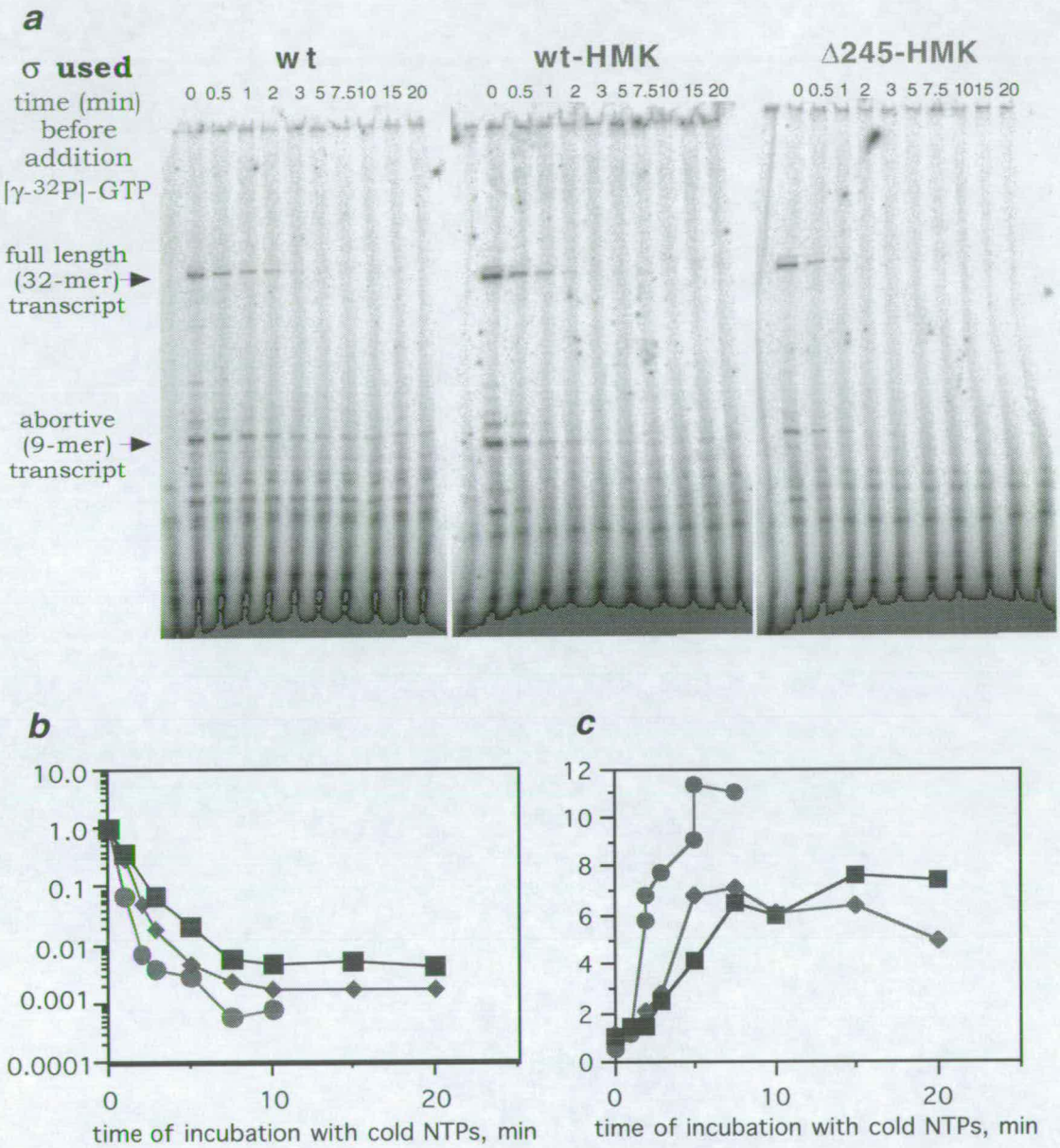


Figure 3.11. Pulse chase assays of E·σ^{70}, E·$\sigma^{70\text{-HMK}}$ and E·$\sigma^{\Delta 245\text{-HMK}}$ (100 nM) on 100 nM λP_R promoter DNA. After the formation of open complex, the transcription was started by adding unlabelled NTPs and heparin, and after indicated time the reaction was chased with the [γ-32P]GTP (see Chapter 2). The reaction products were separated by 20% Urea–PAGE, visualised and quantified. *a)* Gel images of the reactions for the different holoenzymes. The reaction substrates (without DNA and RNA polymerase) are run in the first tracks of every gel. *b)* Amount of the full length chased transcripts relative to that at 0 minutes after the addition of labelled substrate. *c)* Ratios of amount of the full length transcript to that of abortive transcript during the pulse chase reactions. σ subunit used: —■—, wt (σ^{70}); —◆—, wt-HMK ($\sigma^{70\text{-HMK}}$); —●—, Δ245-HMK ($\sigma^{\Delta 245\text{-HMK}}$).

3.9. Sensitivity of both productive and abortive complexes to heparin.

To address the question of sensitivity of productive and abortive complexes to heparin, and, therefore, of stability of these complexes, they were challenged with heparin for different times before the reaction of transcription was started by the addition of NTPs (Fig. 3.12). The results indicate that the reaction of formation both $E{\cdot}\sigma^{\Delta 245\text{-HMK}}_{abort}{\cdot}P$ and $E{\cdot}\sigma^{\Delta 245\text{-HMK}}_{prod}{\cdot}P$ is more reversible than those for $E{\cdot}\sigma^{70}_{abort}{\cdot}P$ and $E{\cdot}\sigma^{70}_{prod}{\cdot}P$, and $E{\cdot}\sigma^{70\text{-HMK}}_{abort}{\cdot}P$ and $E{\cdot}\sigma^{70\text{-HMK}}_{prod}{\cdot}P$.

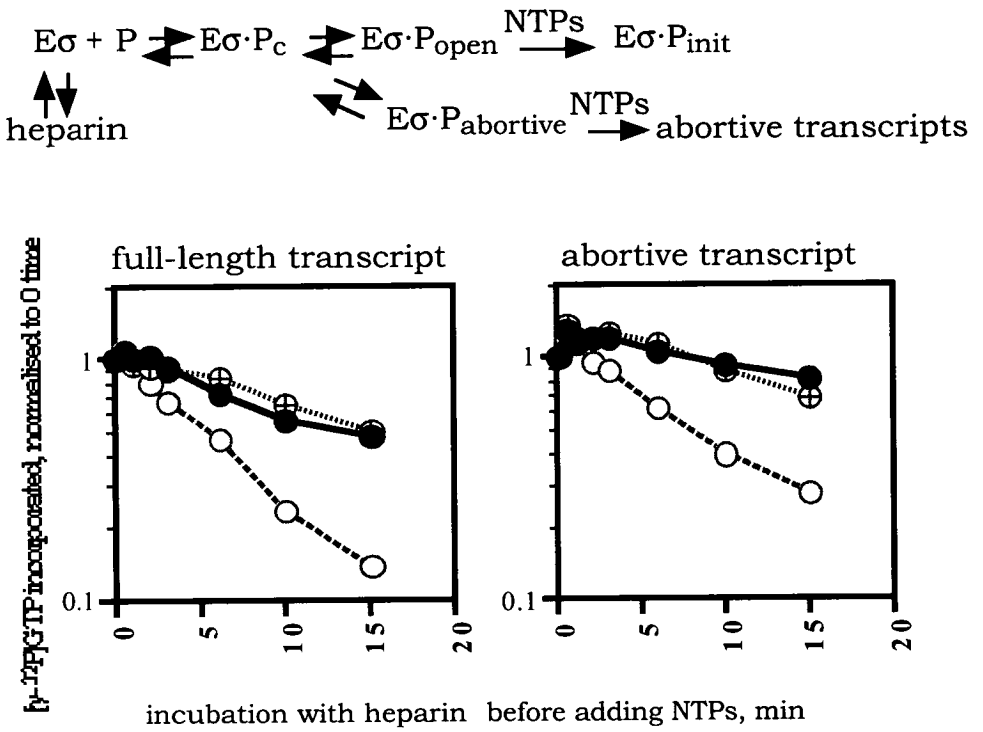


Figure 3.12. Heparin sensitivity of productive and abortive transcription complexes of RNA polymerases with λP_R promoter to heparin. 100 nM λP_R promoter DNA was incubated with 100 nM holo-RNAP for 15 min to form open complex. Then heparin was added to concentration 40 mg ml^{-1}, and after 0 to 15 min incubation the transcription was started by addition of NTPs as described in 2.6.2. The reaction was stopped after 20 min, the reaction products were separated by 20% urea–PAGE, visualised and quantified.

Discussion.

The purpose of this project was to study several chosen amino acids in conserved regions 2 and 3 of σ^{70} and analyse their positions relative to promoter DNA, other regions of σ^{70} and other subunits of RNA polymerase during the binding of sigma to core enzyme, and the forming of closed and open complexes with promoter DNA. The proposed technique suggested using a cysteine-free sigma and substituting the amino acids of interest with cysteine residues. A cysteine-free version of σ^{70}, $\sigma^{\Delta 245}$, earlier shown to be a functional enzyme, was chosen for the study. However, it appeared that $\sigma^{\Delta 245}$ had properties which were different from those of σ^{70}, and therefore it transpired that it was not entirely suitable for studying the functional mechanisms of σ^{70}. In the meantime several publications have appeared where this technique was successfully used in σ^{70} where all three natural cysteines were substituted with alanines [Owens *et al.*, 1998a; Owens *et al.*, 1998b; Bown *et al.*, 1999; Traviglia *et al.*, 1999]. In this work we concentrated on the effects of the deletion of 245 amino acids between conserved regions 1.2 and 2.1 in σ^{70}, and the importance of several amino acids in conserved regions 2.3, 2.4, 3.1 and 3.2 on the function of the subunit.

σ^{70} and $\sigma^{70\text{-HMK}}$ of *E. coli* and σ^{A} of *B. subtilis* are soluble and predominantly exist in a monomeric form in aqueous buffers, including 0.2 M NaCl or 0.5 M ammonium sulphate in TGEB, which is in contrast to hydrophobic nature of $\sigma^{\Delta 245\text{-HMK}}$. They also do not bind hydrophobic chromatography columns (Phenyl Toyopearl) irreversibly, and do not precipitate during *in vitro* refolding when their solution in 6 M guanidine chloride in TGEB is dialysed against TGEB buffer or 0.5 M ammonium sulphate in TGEB [Burgess, 1996; Nagai and Shimamoto, 1997]. σ^{70} is known to have some hydrophobic properties. For example, upon its purification there is always a portion of protein in an oligomeric form, together with a monomeric form, which has to be purified [Burgess, 1996]. The monomeric form of σ^{70} is stable in aqueous buffers at 4°C. The equilibrium between monomeric and oligomeric forms is, however,

shifted significantly towards the oligomeric form for $\sigma^{\Delta 245\text{-HMK}}$, as has been shown in this work. σ^{A}, the major σ factor of *B. subtilis*, does not have a spacer domain between its conserved regions 1.2 and 2.1, but is a soluble monomeric protein *in vitro*.

Lowe *et al.* [1981] showed that σ^{70} forms filaments *in vitro* when heated to temperatures above 49°C. σ^{70} subunit isolated from *rpoD800* (a temperature sensitive allele) strain of *E. coli* which carries a deletion of aa 330–343 in the Δ245 domain, is capable of forming filaments when incubated at temperatures above 37°C *in vitro*. Increasing glycerol or NaCl concentrations reduces the ability of *rpoD800* σ to filament. Filamentous *rpoD800* σ^{70} is inactive, but denaturation of these filaments in 6 M guanidine chloride and then refolding leads to restoration of σ activity. It was suggested that filamentation is a possible route to inactivating σ^{70} under heat-shock conditions, by preventing σ^{70} from interacting with the core enzyme and thus allowing the heat-shock specific σ^{32} to interact with core RNA polymerase and transcribe the genes of the heat-shock regulon. An evidence that the deletion of aa 330–343 (*rpoD800* σ^{70}) changes properties of σ^{70} is that *rpoD800* σ^{70} becomes accessible to attack by the La protease under the heat shock conditions [Grossman *et al.*, 1983]. It can be that the protease recognises a filamentous form of *rpoD800* σ^{70}, or a transitional (from active to inactive) form of it.

Recent work in N. Shimamoto's laboratory has demonstrated that $\sigma^{\Delta 245\text{-HMK}}$ oligomers are also filamentous and, therefore, one possibility is that the deletion of the 245 residues leads to the exposure of two hydrophobic sites in the rest of σ, which causes its oligomerisation by filamentation. It is also possible that the 245 residue domain prevents σ^{70} from filamentation at normal growth temperatures *in vivo*. The deletion present in *rpoD800* σ^{70} is only a small part of the Δ245 region; $\sigma^{\Delta 245}$, however, seems to be much more hydrophobic than *rpoD800* σ^{70} which could be easily refolded from 6 M guanidine chloride solutions by dilution, and is eluted from a gel filtration column as a monomer [Lowe *et al.*, 1981]. This

suggests that further deletion of the amino acids in Δ245 region increases overall hydrophobicity of σ^{70}.

In the X-ray crystal structure of σ^{70} fragment (residues 114–442) [Malhotra *et al.*, 1996] amino acids 330–343 (not present in *rpoD800* σ^{70}) form a loop between helices 10 and 11, and a part of helix 11. This part of the 245 aa domain may be contacting one of the hydrophobic patches elsewhere in σ^{70} outwith the crystallised fragment, thus normally preventing it from exposure to solvent.

From comparison of hydrophobicity analyses of $\sigma^{\Delta 245}$ and σ^{A} [Kyte and Doolittle, 1982] it is not easy to foresee if $\sigma^{\Delta 245}$ has more pronounced hydrophobicity than σ^{A}, and so the reason for the tendency of $\sigma^{\Delta 245}$ to exist in multimeric forms is unclear from the sequence data alone. However, the 245 aa domain in σ^{70} has a very acidic region the role of which is not clear, but presumably this stretch of negatively charged amino acid residues (D and E) influences the relative positions of other σ^{70} domains in the free protein.

As was mentioned in Chapter 1, in the crystal structure of the σ^{70} fragment [Malhotra *et al.*, 1996] the 245 amino acid residue segment between regions 1.2 and 2.1 consists of several α-helices. Therefore by removing these residues from the protein one would expect the overall α-helix content of the protein to be reduced. It also seems likely that this deletion would not be expected to cause a significant change in the structure of the rest of the protein. The reduction of α-helix content was indeed observed by Smillie *et al.* in CD studies of σ^{70} and $\sigma^{\Delta 245}$ (unpublished work by D. A. Smillie, M. K. Kumaran, R. S. Hayward, S. Kelly and N. Price). However, they also observed that the deletion of the 245 residues caused an increase in total β-sheet content in the rest of the protein by at least 23% in the absence of a detergent, which is a radical change in the structure of the σ subunit. In the presence of 0.1% SDS σ^{70} has an increased content of β-sheet, (estimated 115 amino acid residues against 37 in the absence of SDS), closer to that in $\sigma^{\Delta 245}$ both in presence (133 aa) or absence (122 aa) of the detergent. A

possible interpretation of the results was that the 245 residues domain confers a constraint to the structure of σ^{70} thereby preventing some part of it from forming a β-sheet, and that SDS reduces the effect of the 245 residues domain in restraining the structure of σ^{70}. It is possible that, with SDS present in solution, σ^{70} can adopt a favourable β-sheet-rich conformation and expose some of its hydrophobic residues normally inaccessible to solvent in the free σ^{70}.

As was mentioned in Chapter 1, the σ^{70} subunit changes its conformation upon binding to the core RNA polymerase, and further does so when the holoenzyme forms a binary complex with promoter DNA. It has also been shown that some proteins which bind single-stranded DNA have β-sheet-rich structures, and that this mechanism of binding ssDNA is conserved throughout species [Bochkarev *et al.*, 1997]. Therefore it can be proposed that σ^{70} increases its β-sheet content upon formation of an open complex with promoter DNA. It is possible that similar conformational change was observed by Smillie *et al.*, as mentioned above, when detergent was added to σ^{70} in solution, or by deletion of the 245 aa region from σ^{70}.

It has been shown that change in the secondary structure of certain proteins (e.g. human lysozyme; a fragment of yeast protein Ure2p) from α-helical to β-sheet is accompanied by their polymerisation by filamentation [Booth *et al.*, 1997; Taylor *et al.*, 1999].Filamentation of amyloidogenic forms of human lysozyme happens both *in vivo* and *in vitro*, notable that *in vitro* it is temperature induced. The activity of inactive fibrillar lysozyme can be reversed by denaturing and refolding the protein. Mutant lysozyme proteins were shown to have less constrained domain organisation, and they filament more readily. A proposed mechanism for this mode of polymerisation includes transient unfolding of a part of the protein with some α-helical content and then re-folding this part into a β-sheet structure. These β-sheets can then be arranged in layers intermolecularly causing a specific irreversible polymerisation of the protein into filaments. Given the

high β-sheet content of $\sigma^{\Delta 245}$ and its filamentation in solution, it is possible to speculate that its filaments (more specifically those formed by $\sigma^{\Delta 245\text{-HMK}}$) are also generated by intermolecular β-sheet polymerisation. The 245 aa domain therefore can possibly help to keep the overall domain organisation of the rest of σ^{70} more restrained.

The 245 amino acids in σ^{70} between the conserved regions 1.2 and 2.1 are often referred to as "non-conserved", because they are absent from the majority of σ subunits from different organisms, including the major σ subunit of another well studied bacterium, *B. subtilis*. However, this spacer region appears to be conserved among many known major σ subunits of *Proteobacteria*, as can be found by BLAST 2 searches for the 245 aa sequence. These bacteria are: *E. coli, Salmonella typhimurium, S. typhi, S. paratyphi A, Klebsiella pneumoniae, Yersinia pestis, Buchnera aphidicola, Shewanella putrefaciens, S. violacea, Vibrio cholerae, Actinobacillus actinomycetemcomitans, Haemophilus influenzae, Pasteurella multocida, Pseudomonas chlororaphis, P. putida, P. tolaasii, P. fluorescens, P. aeruginosa, P. syringae, Acinetobacter sp., A. lwoffii, Legionella pneumophila, Thiobacillus ferrooxidans, Xanthomonas campestris* (γ subdivision); *Burkholderia cepacia, Bordetella pertussis, B. bronchiseptica, Neisseria gonorrhoeae* (β subdivision); *Caulobacter crescentus, Agrobacterium tumefaciens, Sinorhizobium meliloti, Bradyrhizobium japonicum, Rhodobacter capsulatus, Rickettsia prowazekii* (α subdivision); *Myxococcus xanthus, Campylobacter jejuni* (δ/ε subdivision). As seen from Fig. 3.13, the conservation of the spacer region between regions 1 and 2 of major σ factors is more significant within subdivisions of *Proteobacteria* (marked in brackets after names of the species), and less significant between subdivisions. Even among *Proteobacteria* the spacer region is much less conserved than the highly conserved regions 1, 2, 3 and 4. However, it is always present in the major sigmas of this class of bacteria. Notably, even sigmas of *Proteobacteria* with the least conserved spacer regions (*Myxococcus xanthus, Campylobacter jejuni, Geobacter sulfurreducens*) still have the region of approximately 245 aa between their conserved regions 1 and 2. At the same time

all the known major σ subunits from other groups of bacteria, particularly from gram-positive bacteria *(Bacillus subtilis, B. anthracis, Streptococcus pyogenes, S. pneumoniae, S. pyogenes, Clostridium acetabutylicum, C. difficile, Enterococcus faecalis, Staphylococcus aureus, Corynebacterium diphtheriae, Mycobacterium leprae, M. avium, M. tuberculosis, M. bovis)*, as well as all the alternative sigmas from *Proteobacteria*, do not have or do not seem to have this region. It is possible that under some conditions when the activity of major σ subunit should be reduced (for example heat-shock or in the stationary phase) the 245 domain can help sequester the σ in an inactive form, presumably by polymerisation.

Fig. 3.13. ClustalW alignment of spacer regions of principal sigma factors of proteobacteria (α, β, γ and δ subdivisions). Colour coding: ILVAM (aliphatic/hydrophobic); FWY (aromatic); KRH (positive); DE (negative); STNQ (hydrophilic); PG (conformationally special); C (cysteine). Numbers in brackets are positions of amino acids in the spacer regions of bacteria whose genome sequencing is in progress.

Chapter 4

scrP gene and its product

4.1. Theoretical analysis of ScrP and MtgA polypeptide sequences.

Recently the genomes of several organisms have been completely sequenced, and many more sequencing projects are under way. From the results of a BLAST search [Altschul *et al.*, 1997] of the 23 completely sequenced genomes it is obvious that no amino acid sequences with any significant degree of similarity to ScrP are encoded, except by *E. coli*. These *scrP*-null genomes include those of 6 archaea, 15 bacteria (other than *E. coli*), and 1 eukaryote *(Saccharomyces cerevisiae)*. Among the completely sequenced bacterial genomes, those of species of *Aquificales*, *Firmicutes*, *Spirochaetales*, *Chlamydia*, *Proteobacteria*, and *Cyanobacteria* are represented. However, among the genomes whose sequencing is in progress, several carry ORFs with a significant degree of similarity to *scrP*. Apart from the eukaryotes discussed below, all of these belong to the gamma subdivision of *Proteobacteria (Enterobacteriaceae: Escherichia coli, Salmonella typhimurium, S. typhi, S. paratyphi A, Klebsiella pneumoniae, Yersinia pestis; Vibrionaceae: Vibrio cholerae; Pseudomonas group: Pseudomonas aeruginosa; Shewanella: Shewanella putrefaciens)* or to the *Green sulphur bacteria group (Chlorobium tepidum)* (Fig. 4.1).

As mentioned in chapter 1, significant sequence similarity to ScrP is seen in the potential product of mRNA encoded by human chromosome 21 (KNP-1) and in a protein expressed specifically in fish eye photoreceptor cells (ES1). Moreover, among mRNAs isolated from numerous human, mouse and rat tissues, such as lungs, kidney, heart, mammary gland, myotubes, diaphragm, embryo, placenta, many potentially encode amino acid sequences resembling the KNP-1 protein. It has also worth noting that an mRNA isolated from adult human eye [Hillier and al, 1995] seems to result from primary KNP-1 transcript, different from the one generating the ES1 protein in zebrafish retina (Fig. 4.1).

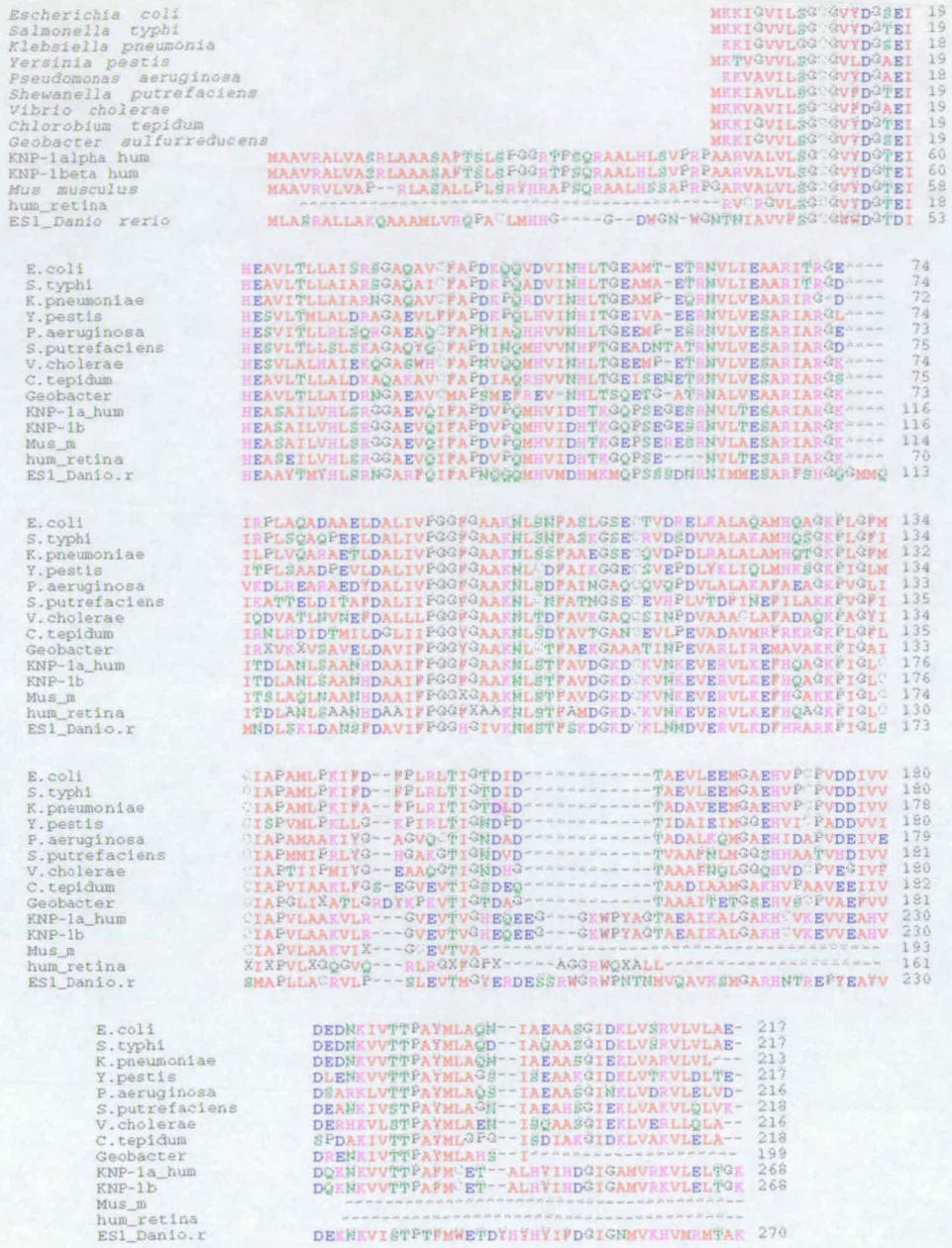


Fig. 4.1. ClustalW alignment of known ScrP analogues, bacterial (from unfinished genome projects), and eukaryotic (KNP-1alpha and KNP-1beta, different splicing forms of human KNP; ES1 – protein from zebrafish eye; EST tag from human retina of an adult [Hillier *et al.*, 1995]). Colour coding: ILVAM (aliphatic/hydrophobic); FWY (aromatic); KRH (positive); DE (negative); STNQ (hydrophilic); PG (conformationally special); C (cysteine).

Recently submitted sequences of regions of human chromosomes 7, 14 and 16 contain more ORFs which can potentially encode polypeptides similar to parts of ScrP [GenBank Accession Numbers AC009196, and AC009034, EMB Accession Number AC008015]. This is consistent with the earlier suggestion [Nagamine *et al.*, 1996] that more than one protein similar to ScrP may be produced in vertebrates, with different tissue or developmental specialisation. However, too high degree of similarity of these potential polypeptides to ScrP of *E. coli* (nearly 100% for polypeptides of 100–120 aa long) may also indicate of cloning error.

In any case, proteins similar to ScrP are present in many member species of the gamma subdivision of *Proteobacteria* and in at least one *Green Sulphur bacterium* (Fig.4.1), but absent from many other classes of bacteria and even from *Haemophilus influenzae*, which belongs to the *Pasteurellaceae* family of the gamma subdivision of *Proteobacteria*. The conservation among ScrP analogues from different species is strikingly high. Among bacteria, 73–94% amino acids in ScrP analogues are similar, and 57–88% are identical to ScrP. In eukaryotic proteins there is 46% (KNP-1alpha) and 32% (ES1) of identity to ScrP, and 64% and 59% of similarity, respectively. This similarity continues throughout the whole length of the proteins, with the exception of *N*-termini, which are presumably signal peptides. As it was suggested [Scott *et al.*, 1997], such a high degree of conservation may indicate that ScrP and its analogues are involved in a basic molecular mechanism common to bacteria and higher organisms. Unexpectedly, proteins with any degree of similarity to ScrP are absent from *Saccharomyces cerevisiae*, but present in higher eukaryotes including fish and mammals. These findings suggest that the physiological functions of proteins related to ScrP are important, but cast no light on their actual functions.

Interestingly, in the *Enterobacteriaceae* the genes potentially coding for ScrPs are present in the same environment as that of *scrP* on the *E. coli* chromosome, i.e. the downstream ends of *scrP* in *E. coli*, *Salmonella typhi*, *S. typhimurium*, *S. paratyphi A*, *Klebsiella*

pneumoniae and *Yersinia pestis* all overlap 1–4 bp with the upstream ends of genes similar to *mtgA*, and a gene similar to *arcB* is present about 230 bp upstream from the start codon of each *scrP*. This could either be due to merely their evolutionary proximity, or the arrangement might be conserved because the functions of ArcB, ScrP and MtgA are connected. However, in the other bacterial genomes possessing a gene similar to *scrP* but belonging to different taxonomic groups in the gamma subdivision of *Proteobacteria (Vibrio cholerae, Pseudomonas aeruginosa, Shewanella putrefaciens)* or to *Chlorobium tepidum (Green Sulphur Bacteria group)* (Fig. 4.1), the *scrP* analogue is not accompanied by genes similar to *mtgA* or *arcB*, thus undermining the proposal that their proximity in the *Enterobacteriaceae* implies related functions.

Searches for a signal peptide in ScrP using the computer programmes SignalP and PSORT (see chapter 2.7) produce contradictory results. One SignalP prediction, based on the Hidden Markov model, suggests that ScrP is unlikely to contain a cleavable signal peptide and thus is not a secreted protein. However, the other SignalP prediction, based on the Neutral Networks model, suggests that ScrP might be cleaved between amino acids 36 and 37.

PSORT analysis also predicts that ScrP is a secreted protein, with a possible signal peptide cleaved between amino acids 34 and 35. Moreover, it predicts that ScrP should be localised in the periplasmic space, with a probability of 0.92.

The human ScrP-like protein KNP-1 has similarly been found to have a potential signal peptide for targeting it into mitochondria, while ES1, expressed specifically in adult zebrafish eye photoreceptor cells, has been shown to be neither a nuclear nor a membrane-associated protein. PSORT2 and SignalP analyses of KNP-1 show a 65% likelihood that it is a mitochondrial polypeptide. As for ES1, the predictions for localisation are: cytoplasmic 40%, nuclear 30%, mitochondrial 17%. As already mentioned above and

in chapter 1, ES1 is not in fact a nuclear protein, but it could be a cytoplasmic protein.

Similar analyses were carried out for MtgA. All the computer programmes used predict the presence of one transmembrane region in MtgA, between amino acids 15–21 and 36–39. Furthermore, according to TMHMM, TMpred, PHDhtm, and HMMTOP (see chapter 2.7), the *N*-terminus of MtgA should be oriented into the cytoplasm, and its *C*-terminus into the periplasm.

4.2. *In vivo* expression of the plasmid-borne genes of the *arcB* operon.

4.2.1. Expression of *scrP* and *mtgA* in mini- and maxicells.

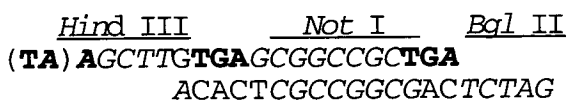
D. Smillie and F. Townsley conducted an analysis of the proteins encoded by the *Hin*d III–*Eco*R I insert of pFMT1 in minicells [Smillie, 1994]. They were able to identify polypeptides of estimated molecular size ~14 kDa and ~28 kDa on SDS–PAGE, and the 28 kDa polypeptide was thought to be ScrP. However, the *Hin*d III–*Eco*R I insert in pFMT1 has the potential to encode at least three polypeptides – ArcB, ScrP and the hypothetical MtgA, with calculated sizes of 88 kDa, 23 kDa and 27 kDa respectively. The 28 kDa band migrates between the precursor and mature forms of plasmid encoded β–lactamase, which makes it difficult to judge to which of the proteins – ScrP or MtgA – it actually corresponds. The presence of the second of these proteins could have been masked by the β–lactamase bands. Moreover, in their experiments only a faint protein band possibly corresponding to ArcB can be seen. At the time it seemed possible that pFMT1 might generate a 5'-truncated *arcB* mRNA, leading to possible internally-started translation. We now know that pFMT1 has a full-length *arcB* gene with the ribosome-binding site changed by *EcoR* I cleavage. The 14 kDa polypeptide also remained unexplained.

To further investigate proteins encoded by the *Hin*d III–*Eco*R I region, several other plasmids with disrupted *scrP* and *mtgA* genes were constructed and analysed, before and during the present work.

D. Smillie inserted a Kan^R cassette into the *Nsi* I site of *scrP* in pFMT1, to produce pDSA28 [Smillie, 1994]. The orientation of the Kan^R cassette is opposite to that of the *scrP* gene, which was confirmed by restriction digests with *Hind* III, *Nru* I, and *Xho* I (Fig. 4.2).

pFMT1 has the λ522 clone of the Kohara *E. coli* library [Kohara *et al.*, 1987] (see Fig 1.2.1) as its source of genomic DNA. In this clone, the beginning of *arcB* is situated very near the end of the chromosomal insert, and it was thought possible that *arcB* might be incomplete Accordingly A. Elding in this laboratory constructed a new plasmid, pACE1, using the 5.34 kb *Bam*H I–*Eco*R I fragment from the λ523 clone of the Kohara library [Elding, 1995]. *arcB* is situated in the middle of the λ523 DNA. However, the *Eco*R I recognition sequence lies in the ribosome-binding site of the gene, so that even though pACE1 (and probably pFMT1) carry a full-length *arcB* coding sequence, the original ribosome-binding site is damaged (Fig. 4.2).

pACE1 carries a large amount of DNA downstream of the *Hind* III site (which conveniently overlaps the stop codon of *mtgA*). In order to remove it the *Hin*d III–*Bam*H I chromosomal fragment from pACE1 was replaced with a synthetic *Hin*d III – *Not* I – *Bgl* II linker (primers TK1 and TK2), thus destroying the *Bam*H I and introducing a *Not* I site together with stop codons in all three frames in the direction of *arcB*, *scrP* and *mtgA* translation:



The resulting plasmid was called pTAK1, and its structure was verified by restriction analysis with *Hin*d III, *Bam*H I, and *Not* I (Fig. 4.2).

The minicell-producing strain P678.54 was separately transformed with the plasmids – pJJW11, pACE1, pFMT1, pTAK1, pDSA28 and pBluescript::Kan. Minicells were prepared and plasmid-encoded proteins labelled with [^{35}S]Met as described in sections 2.3.7, 2.3.8. The labelled proteins in the crude minicell extracts were separated by 10% SDS–PAGE and visualised by autoradiography (Fig. 4.3).

It can be seen that pTAK1 produces essentially the same pattern of expressed proteins as pFMT1. Both plasmids, as well as pACE1, encode a protein of ~29 kDa, which appears on SDS–PAGE as a band between the two β–lactamase bands; on the other hand, cells carrying pDSA28, in which *scrP* is disrupted with the Kan^R cassette, do not produce this protein. This suggests that the protein corresponding to this band is either ScrP or MtgA. The absence of the latter from pDSA28 track could be either due to interruption of *mtgA* transcription by the insert in *scrP*, or due to obligatory coupled translation of *mtgA* with *scrP*. However, the results obtained with maxicells (discussed later) indicate that the MtgA runs as ~25 kDa protein in my gels. Accordingly the 29 kDa band in Fig. 4.3 is very likely to correspond to ScrP.

The presence of the ~14 kDa product is also evident in several tracks. This will be discussed later.

In order to learn more about the identities of these proteins and about the transcription of the region, new disruptions were made (Fig. 4.4 and Table 4.1).

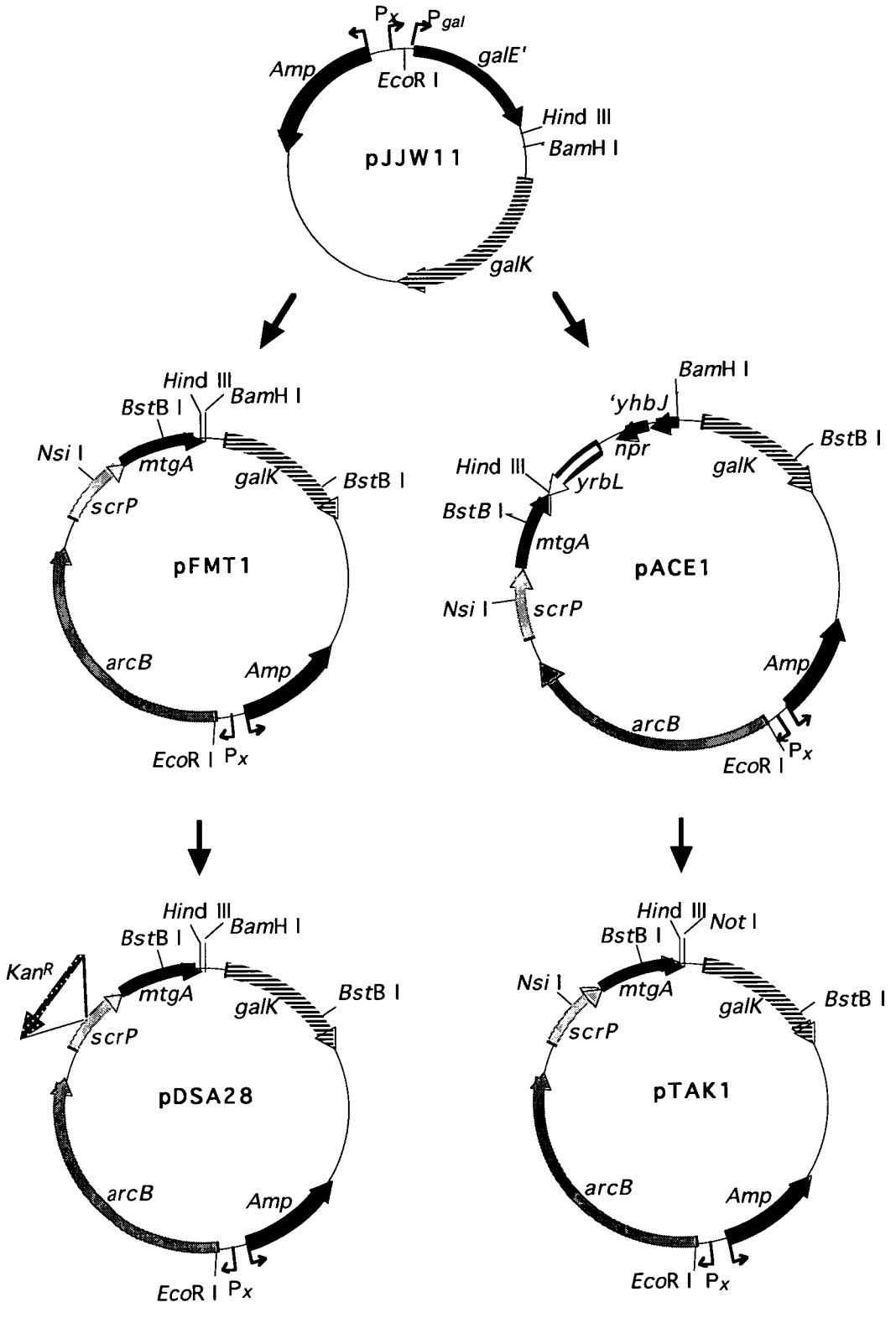


Figure 4.2. Scheme of the construction of plasmids pFMT1 [Townsley, 1991], pDSA28 [Smillie, 1994), pACE1 [Elding, 1995], and pTAK1 (this work). The genes, relevant restriction sites and known promoters are shown. *galK* is promoterless. P_x – a plasmid promoter (pJJW11).

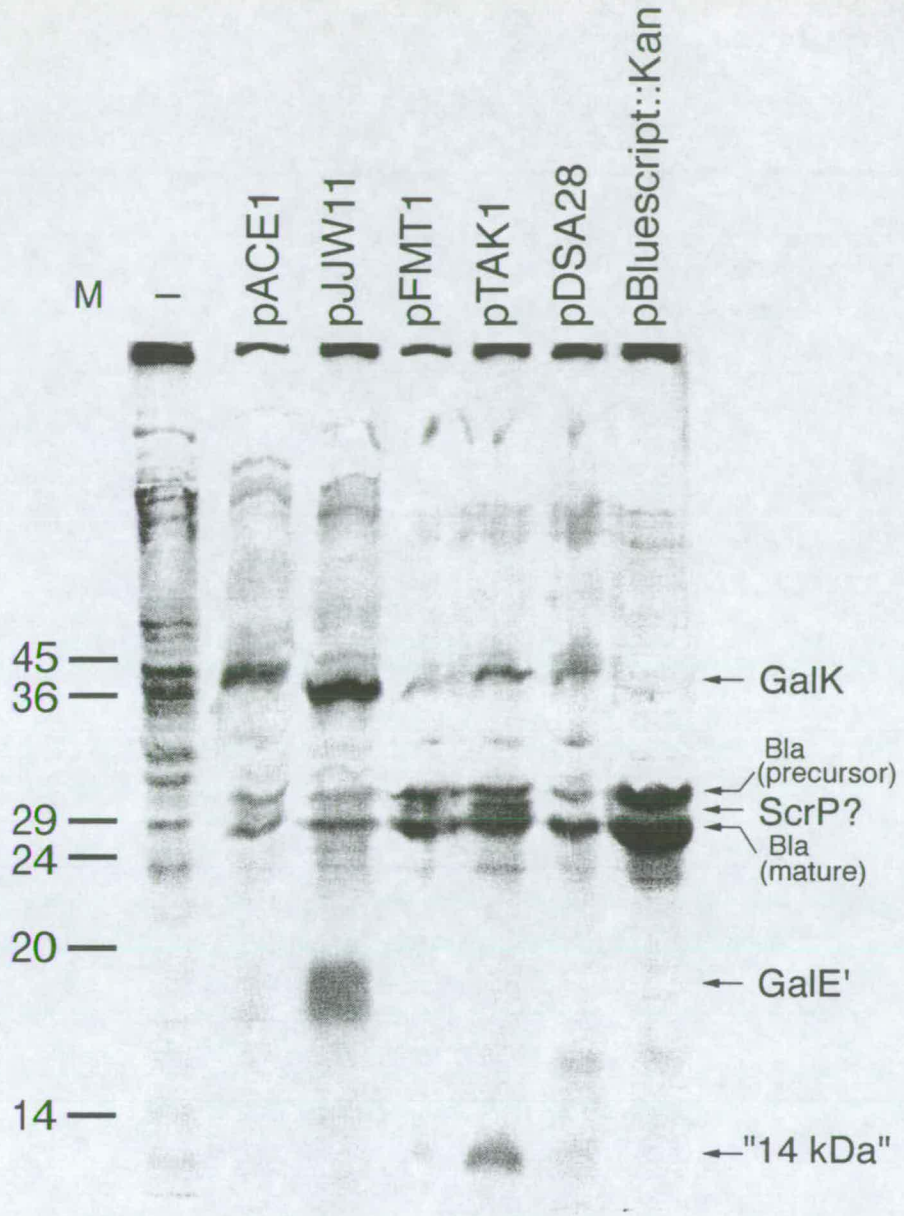


Figure 4.3. Plasmid-encoded proteins expressed in minicells. *M* – molecular size marker, kDa; P678.54 cells carrying various plasmids; "–" control experiment, P678.54 cells with no plasmid. Bands corresponding to GalK, precursor and mature β-lactamase, GalE', 14 kDa protein and, presumably, ScrP, are shown on the right. The difference in GalK, ScrP and 14 kDa protein relative band intensities in pACE1, pFMT1, and pTAK1 are presumably an artefact due to different amount of loaded material in the tracks. pBluescript::Kan was used as a control for β-lactamase and Kan^R.

pTAK1 was partially digested with *Bst*B I and ligated with a 1.25 kb *Acc* I fragment from pUC4K carrying a Kan^R cassette. DNA restriction with these enzymes produces compatible cohesive ends. Several ligation products were produced: the insertion of the Kan^R

cassette into the *Bst*B I site in *mtgA* in either orientation, resulting in pTAK2 and pTAK3; the replacement of the *Bst*B I fragment of pTAK1 with the Kan^R cassette in either orientations, yielding in pTAK4 and pTAK5; and the insertion of the Kan^R cassette into the *Bst*B I site in *galK* in the same orientation as the *galK* gene, resulting in pTAK6 (Fig. 4.4.). The structures of these plasmids were determined by restriction analysis with *Nru* I, *Eco*R I, *Bst*B I, *Nsi* I.

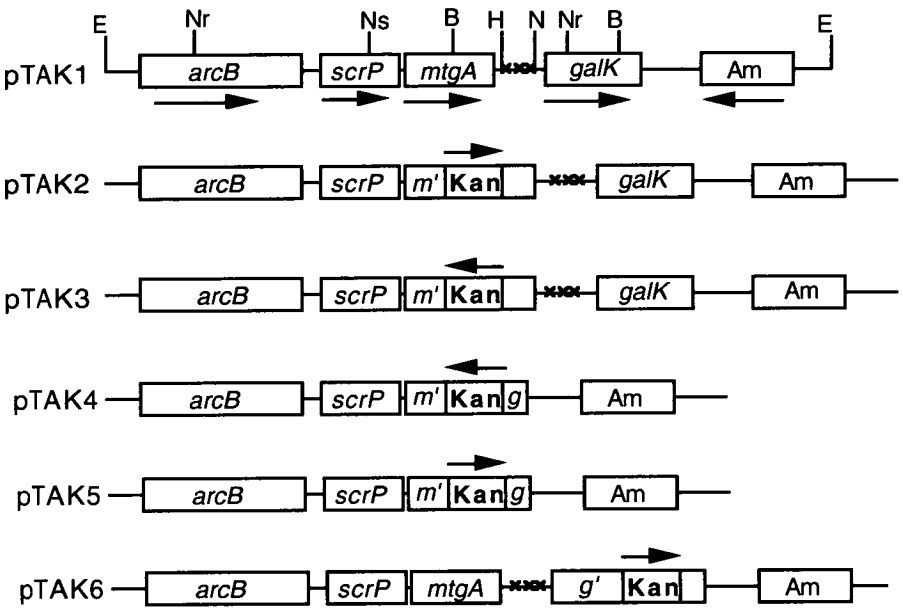


Figure 4.4. Plasmids constructed for studying expression of *arcB–scrP* region. Orientation of genes and relevant restriction sites are shown. *m'* – *mtgA'*, *g'* – *'galK*, E – *Eco*R I, Nr – *Nru* I, Ns – *Nsi* I, B – *Bst*B I, H – *Hin*d III, N – *Not* I. xxx – translation stop codons in all three frames in the direction from *mtgA* to *galK*. *Nru* I, *Eco*R I, *Bst*B I, *Nsi* I

Table 4.1. Orientation and calculated sizes of the products of some chimaeric genes presented on plasmids used in the maxicell experiment.

Plasmid	Chimaeric genes constructed	Orientation of Kan^R relative to *scrP* (→)	Calculated product sizes
pDSA28	*scrP'* + tail	←	16 kDa
pTAK3, 4	*mtgA'* + tail	←	16.2 kDa
pTAK2, 5	*mtgA'* + tail	→	18.7 kDa

The *E. coli* strain CSR603, designed for producing maxicells, was transformed with all of the above plasmids. Maxicells were then

prepared according to the protocol described in sections 2.3.9. The proteins synthesised were labelled with [^{35}S]Met. The crude maxicell extracts were separated by 12% SDS-PAGE and the labelled proteins were visualised by autoradiography.

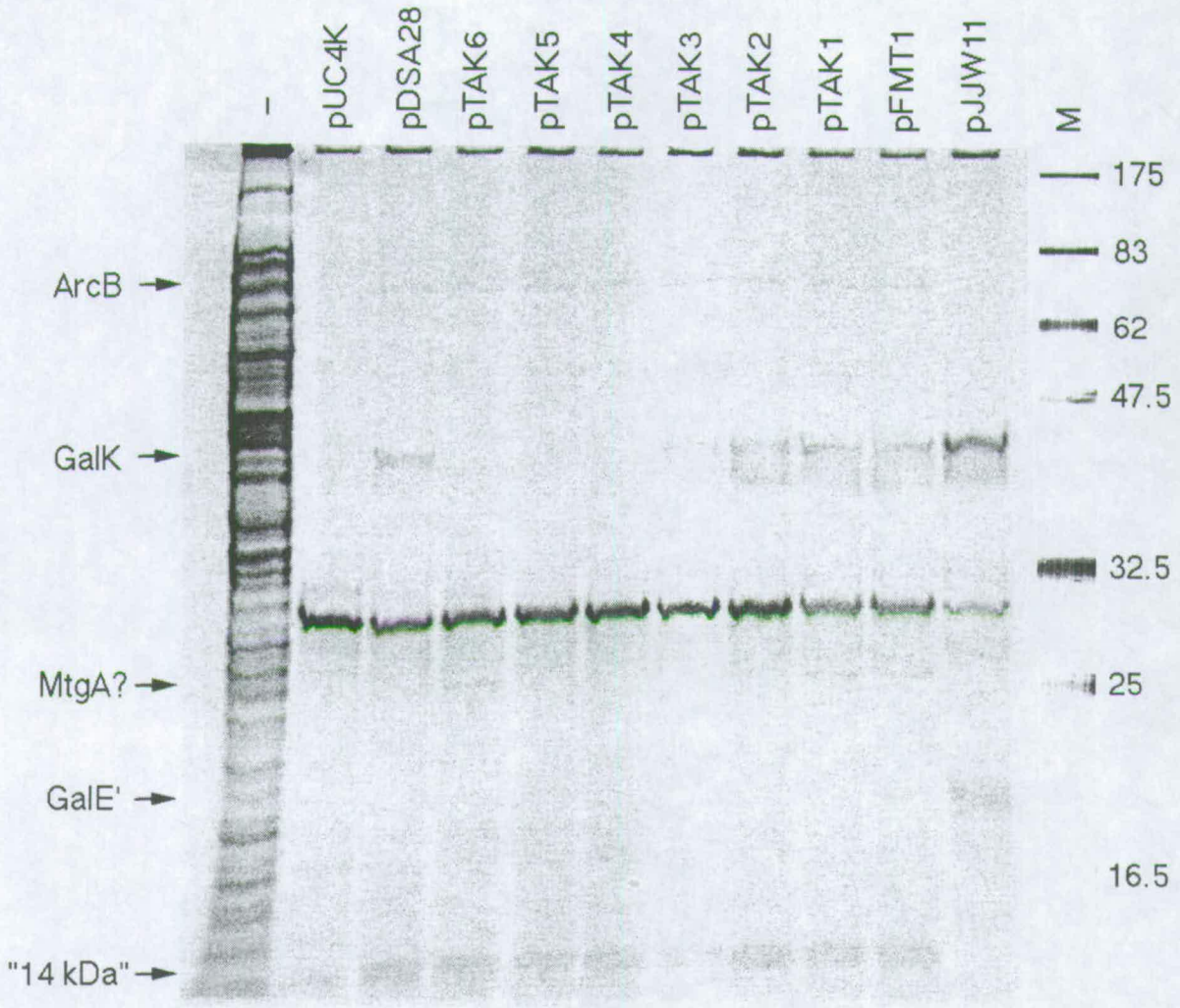


Figure 4.5. Plasmid-encoded proteins expressed in maxicells.
M – molecular size marker, kDa; CSR603 cells carrying various plasmids; "–" control experiment, CSR603 cells with no plasmid. Bands corresponding to ArcB, GalK, β lactamase, GalE', 14 kDa protein and, presumably, MtgA, are shown on the left. pUC4K was used as a control for β-lactamase and Kan^R.

In this experiment a faint band of ~70 kDa appeared in the tracks corresponding to pFMT1, pDSA28, and pTAK1 – pTAK6, but not in the tracks containing the control pUC4K or pJJW11. From earlier work [Iuchi *et al.*, 1990; Iuchi and Lin, 1992a] it is known that ArcB

migrates between 72 and 82 kDa on SDS–PAGE, so this band presumably corresponds to ArcB. As seen from the figure, cells carrying pFMT1 as well as pTAK1 produce this protein. This is consistent with our expectation that pFMT1 has a full-length *arcB* gene, contrary to the uncertainty on this point which led to the construction of ArcB.

The apparent presence of the comparable level of ArcB protein in pDSA28 suggests that *arcB* transcription has not been significantly affected by the insertion of an oppositely oriented Kan^R cassette within *scrP* (Fig. 4.5). This suggests that the cassette – or the DNA between *Nsi* I site of *scrP* and the end of *arcB* – contains a terminator strong enough to prevent much convergent transcription into *arcB*.

It is noteworthy that the 14 kDa protein band is seen in all tracks except the controls (pJJW11 and pUC4K) and at much the same level, irrespective of Kan^R disruptions of *scrP*, *mtgA*, or *galK*. This supports the hypothesis (discussed later) that this in fact is an internal, out-of-frame translation product of *arcB* mRNA.

The band of ~29 kDa, characteristic of one of the proteins encoded by the *Hin*d III–*Eco*R I inserts in pFMT1, pTAK1 and pACE1 and observed in the minicells experiments, was not detected in the maxicell experiment, presumably because it was not resolved from the strong and apparently superimposed β–lactamase and KanR (calculated molecular size 30.7 kDa) bands. However, a new faint band of apparent molecular size ~25 kDa was produced by the cells carrying pFMT1, pTAK1 and pTAK6, where both the *scrP* and *mtgA* genes are intact. This band was absent from pTAK2 – pTAK5, pDSA28, and the control pUC4K and pJJW11 tracks.

Both pTAK2 and pTAK5 have *scrP* intact and *mtgA* disrupted at *Bst*B I by the Kan^R cassette, with the Kan^R gene being in the same orientation as *mtgA*. They should produce a truncated MtgA' with a chimaeric tail, of total size 18.7 kDa. pTAK4 and pTAK3 have *scrP* intact and *mtgA* disrupted by the Kan^R cassette in the orientation

opposite to *mtgA*, and should produce a chimaeric MtgA' of 16.2 kDa. There are several faint bands in this size range, but these are common to all the samples. Perhaps the chimaeric proteins are produced but unstable, or their are masked by the other common bands.

Although pDSA28 has *scrP* disrupted by the Kan^R cassette, *mtgA* is intact. However, the absence of the MtgA (~25 kDa) band from pDSA28 track is understandable if *mtgA* is not transcribed in this plasmid (because transcription from *arcB* and *scrP'* is blocked when it enters the Kan^R cassette) and/or if translation of *mtgA* is obligatorily coupled to that of *scrP*. Accordingly it seems reasonable to conclude that the ~25 kDa band seen only in the pFMT1, pTAK1 and pTAK6 tracks is formed by MtgA protein.

In maxicells carrying plasmids pFMT1, pTAK1, pTAK2, pTAK3 and pDSA28 the full-length GalK polypeptide of calculated molecular size 41.3 kDa is synthesised, and the intensity of the corresponding band has been measured by densitometry. In pJJW11 the *galK* gene is known to be transcribed from its natural promoter, P_{gal} [Wright and Hayward, 1987]. This is absent from pFMT1 and pTAK1, so *galK* transcription becomes dependent on the plasmid promoter P_x [Wright and Hayward, 1987], just upstream from the *Eco*R I site (see Fig. 4.2), and any promoters present in the DNA insert in pJJW11. These plasmids generate about 4-fold less GalK than pJJW11.

pTAK2 produces a surprisingly similar amount of GalK, considering that Kan^R cassette is inserted within *mtgA*. This can be due to possible termination which balances out the transcription from Kan^R and P_x promoters by the terminator whose likely presence in the Kan^R cassette has been mentioned.

pTAK3 and pDSA28 generate ~7 fold less galactokinase than pJJW11, and in these cases *galK* mRNA synthesis is due to transcription coming either from P_x hindered by transcription from the Kan^R promoter in the opposite direction, or from the

promoter(s) lying within the portion of *mtgA* downstream of its *Bst*B I site or in the Kan^R cassette oriented in the direction opposite to the Kan^R gene.

In pTAK4 - pTAK6, the *galK* gene is either disrupted or 5'-truncated, so the failure to produce GalK is expected.

4.2.2. Overexpression of *scrP* and *mtgA* using a T7 RNA polymerase-driven promoter.

scrP and *mtgA* were cloned into pT7-7 [Tabor and Richardson, 1985] under the control of its T7 RNA polymerase-driven promoter, so that upon T7 RNA polymerase induction the ScrP and MtgA polypeptides should be overproduced. Plasmids were constructed in which *scrP* together with *mtgA*, or *mtgA* alone, can be transcribed from the T7 Ø10 promoter.

The *Mfe* I–*Hin*d III fragment of pTAK1 was cloned into the pT7-7 vector digested with *Eco*R I and *Hin*d III, to produce pTAK9. Similarly *Vsp* I–*Hin*d III and *Bfa* I–*Hin*d III fragments of pTAK1 were cloned into pT7-7 digested with *Nde* I and *Hin*d III, to produce pTAK10 and pTAK11 respectively. Thus three different plasmids were obtained in which *scrP* and *mtgA* are fused to the T7 Ø10 promoter.

To clone *mtgA* without intact *scrP*, the *Nsi* I–*Hin*d III fragment of pTAK1 was cloned into pT7-7 digested with *Pst* I and *Hin*d III, yielding pTAK8. Similarly the *Xmn* I–*Hin*d III fragment of pTAK1 was cloned into pT7-7 digested with *Sma* I and *Hin*d III to produce pTAK7.

The structures of all the plasmids shown in Fig. 4.6 were verified by restriction analysis with *Hin*d III and *Cla* I.

Subtle differences exist between the forms of ScrP and of MtgA expected to be produced by these plasmids.

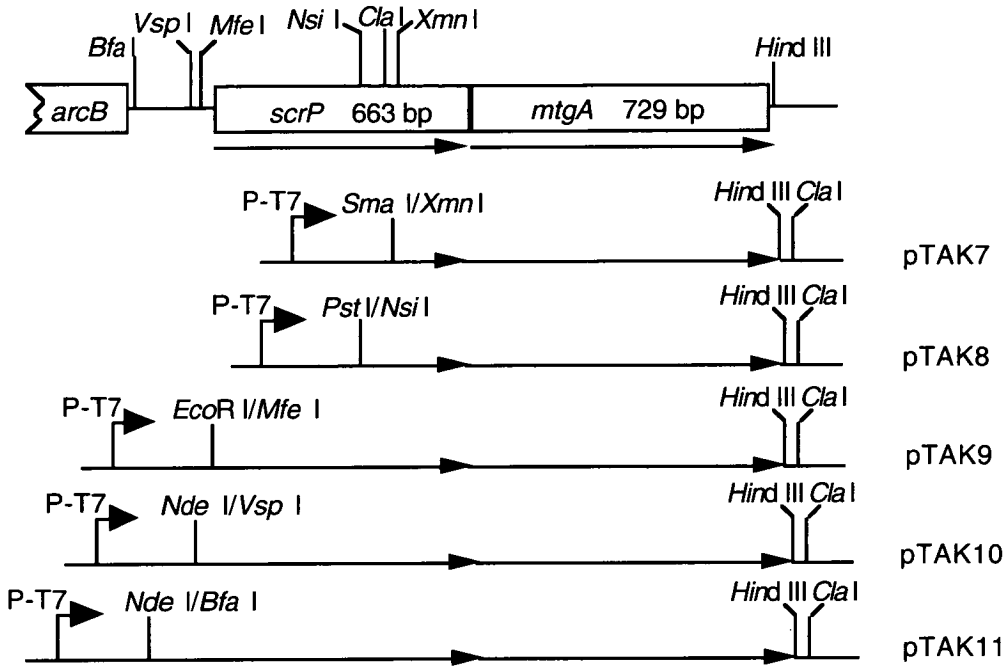


Figure 4.6. Cloning of *mtgA* and *scrP* into pT7-7. Relevant restriction sites are shown.

In the case of pTAK9 (see below) ribosomes initiating at the T7 gene 10 ribosome-binding site, would start at AUG codon in the *Nde* I site producing a heptapeptide terminated by the UGA stop-codon in the ribosome-binding site of *scrP*. They are then likely to reinitiate either at the AUG codon of *scrP*, or at an AUG codon within the *scrP* ribosome-binding site. In the latter case the ScrP polypeptide produced will have three extra amino acid residues (MIT) at the *N*-terminus:

pTAK9

```
     gene 10 rbs      Nde I      Eco RI/Mfe I     scrP rbs
PØ10..AGGAGAT ATACAT ATG GCT AGA ATT GCG AAA GGATGATC
                      M   A   R   I   A   K   G  *
                                                 M  I

    ScrP start
ACA ATG AAG AAA...
T   M   K   K...
```

In the case of pTAK10 and pTAK11, the *Nde I* site and the start codon of the T7 gene 10 are destroyed. Thus in the case of pTAK11 translation is likely to start at the natural *scrP* start codon, with ribosomes binding at the *scrP* ribosome binding site.

For pTAK10, however, there are two possibilities illustrated below. One is that ribosomes would start translation at the AUG codon immediately after the destroyed *Nde* I/*Vsp* I site, leading to the production of a 25 amino acid polypeptide, and would then be likely to reinitiate in another frame, producing another 10 amino acid polypeptide. This process would interfere with the translation of ScrP. Another possibility is that the ribosomes would initiate at a GUG codon 2 nucleotides downstream of the above AUG, and the resulting octapeptide would terminate at the *scrP* ribosome-binding site at the UGA codon described for pTAK9. These ribosomes would then reinitiate either at the natural ribosome-binding site of *scrP* or three codons further upstream.

pTAK10

```
   gene 10 rbs   Nde I/Vsp I                          scrP rbs
PØ10..AGGAGAT ATACA TAATGTGTTTCAGTCAATTGCGAAAGGATGATCACA
                    M  C  F  S  Q  L  R  K  D  D  H  N
                      V  F  Q  S  I  A  K  G  *
                                              M  I  T

ScrP start
ATGAAGAAAATTGGCGTAATTCTGAGCGGATGCGGCGTCTATGAC...
  E  E  N  W  R  N  S  E  R  M  R  R  L  *
                                        M...
M  K  K  I  G  V  I  L  S  G  C  G  V  Y  D...
```

pTAK11

```
    gene 10 rbs    Nde I/Bfa I
PØ10..AGGAGAT ATACA TAGACCGGGGTGCGCGAATAC...

                                  scrP rbs      ScrP start
...ATTAATGTGTTTCAGTCAATTGCGAAAGGATGATCACAATGAAGAAAATTGGCG
                                              M  K  K  I  G  V

TAATTCTGAGCGGATGCGGCGTCTATGAC...
  I  L  S  G  C  G  V  Y  D...
```

In the case of pTAK7, one would expect production of a protein P7–'ScrP carrying the first seven amino acids encoded by the polylinker of pT7-7 (MARIRAL) fused to amino acids 145 – 217 of ScrP (total calculated molecular size 8.7 kDa), together with MtgA. pTAK8 is similar to pTAK7 except that the fusion protein P8–'ScrP encoded by the polylinker of pT7-7 and a part of ScrP should have a molecular size 11.6 kDa (that is, MARIRARGSSRVDL fused to amino acids 125 – 217 of ScrP). MtgA in pTAK7 and pTAK8 is either translationally coupled to ScrP, or independently translated from its own weak ribosome-binding site, or both – just as with *mtgA* encoded by the bacterial chromosome. However, the expression of *mtgA* relative to that of *scrP* from T7 RNA polymerase promoter can be lower than when it is expressed from *E. coli* RNA polymerase promoter, due to the effect known for "T7-overproduction systems" when two successive independently translated genes are expressed [Iost and Dreyfus, 1995]. mRNA between fast moving T7 RNA polymerase and the lagging behind front ribosome is likely to be subjected to attack by nucleases.

The above plasmids were transformed into the *E. coli* B strain BL21(λDE3)plysS, and protein overexpression and labelling were carried out as described in section 2.5.3. The crude cell extracts were separated on 12% SDS–PAGE, and the labelled proteins were then visualised by autoradiography (Fig. 4.7).

ScrP and MtgA contain 7 methionines each. Assuming that they both retain or lose their initial methionine, this allows estimation of the relative amounts of these proteins being synthesised simply by measuring the intensities of the corresponding bands on the autoradiograph.

The autoradiograph of the gel (Fig. 4.7) reveals two bands of ~27 and ~27.5 kDa, present in all induced cells carrying plasmids with full length *scrP* and *mtgA* genes; these products are absent from the cells carrying plasmids encoding MtgA and a chimaeric truncated ScrP. These strong doublet bands are presumably ScrP. The doublet appearance will be discussed later. The third ScrP band, below the

~27 kDa band, is faint and obvious in pTAK11 track. The identity of ~26 kDa band in the pTAK10 track is not clear.

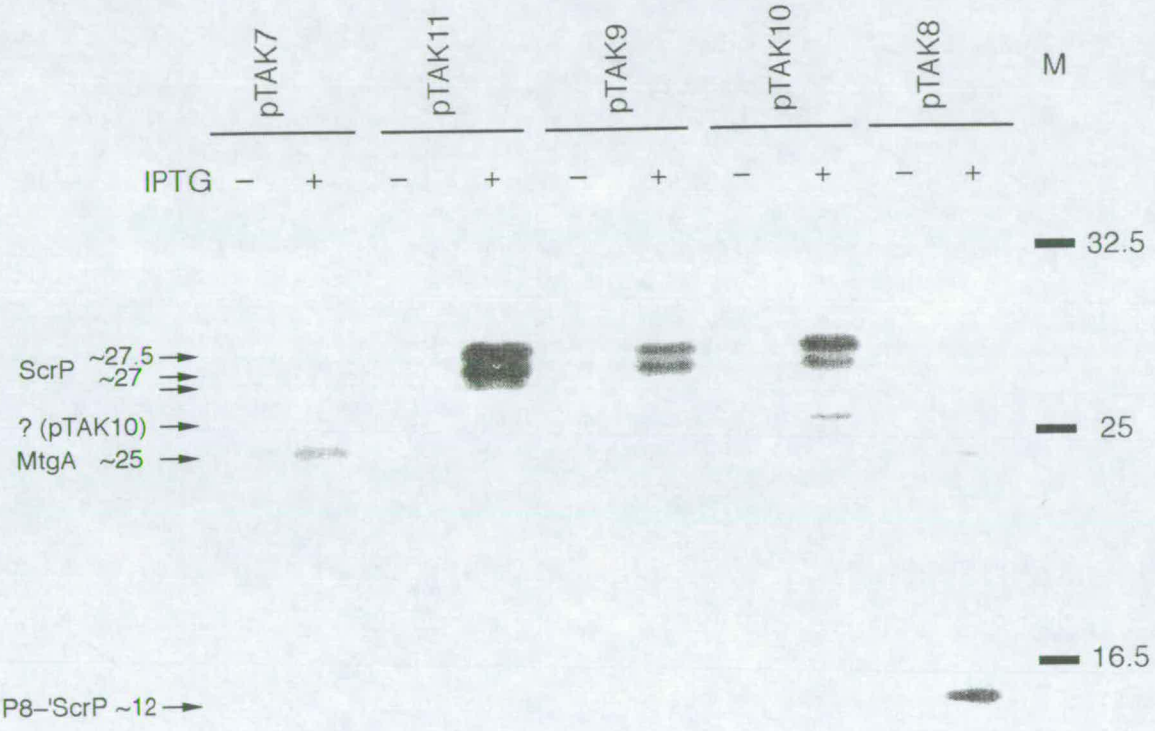


Figure 4.7. Autoradiograph of the gel of the cells overexpressing ScrP and MtgA from pT7-7 derivatives. + or – *IPTG* – with or without T7-RNA polymerase induction. *M* – molecular size marker. Bands sizes and identities are shown on the left. The two bands of ~27 and ~27.5 kDa appearing in the induced cells carrying pTAK9, pTAK10 and pTAK11 should correspond to ScrP.

On the other hand, all five IPTG-induced extracts from cells carrying pTAK7—pTAK11, show a band of ~25 kDa, which presumably corresponds to MtgA. This band is fainter in the case of the plasmids encoding full-length *scrP* upstream of *mtgA*. This is what one would expect if translation of *mtgA* is obligatorily dependant on ribosomes reinitiating after termination of *scrP*, i.e. on translational coupling. In the cases of pTAK7 and pTAK8 the

ribosomes reaching the end of *scrP* will have initiated at the strong T7 gene 10 ribosome-binding site, whereas for pTAK9—11 they will have used the weaker site of *scrP* itself.

As expected, the IPTG-induced cells carrying pTAK8 produce a strong band of ~12 kDa, its intensity being about 5–7 times that of MtgA. The predicted peptide should contain 5 methionines, assuming that the *N*-terminal one is not cleaved off. This suggests that about one MtgA molecule is synthesised per 7–10 P8-'ScrP molecules. Similar calculations for pTAK9, pTAK10 and pTAK11, counting both bands of the doublet and the minor band as ScrP, give ratio of ScrP to MtgA produced 30, 16 and 44, correspondingly, with the error of the measurement about 20%. Because of the unexplained band, which can be a truncated form of ScrP, present in pTAK10 track, the ScrP/MtgA ratio is the lowest for this plasmid. Considering pTAK9 and pTAK11 results, one MtgA molecule is synthesised per every 30–50 molecules of ScrP in this experiment, which is even less than in case of pTAK8 which carries a truncated P8-'ScrP. This can be, apart from the big error of measurement, due to the mentioned above preference of "T7 overproduction" system for the upstream gene. According to this view, *mtgA* should be expressed better in pTAK7 than in pTAK8, and in pTAK8 better than in pTAK9—11, because the distance between ribosome-binding site and the start of *mtgA* is the shortest in case of pTAK7, longer in case of pTAK8, and therefore *mtgA* has the better chance to get expressed than in cases of pTAK9—11.

Consideration of these results suggests quite strongly that *mtgA* translation is rarely initiated independently. Instead it is dependant on reinitiation by ribosomes which have already translated *scrP*, with an efficiency of some 2–10%, at least under the conditions of this *in vivo* experiment.

4.2.3. Cell fractionation.

BL21(λDE3)plysS cells carrying pTAK8 and pTAK11 which had produced the most interpretable results above, were chosen for investigation of intracellular localisation of ScrP and MtgA.

Cells induced and labelled with [^{35}S]Met as above were subjected to fractionation as described in Section 2.5.3. Periplasmic, total membrane, outer membrane and cytoplasmic fractions were prepared. Their proteins were then fractionated on 10% tricine SDS–PAGE and examined by reverse staining (Fig. 4.8) and by autoradiography (Fig. 4.9).

In the case of pTAK8, the overproduced MtgA is evident in the total cell extract, but does not appear in any of the cell fractions (Fig. 4.9). This suggests that the protein has been lost during cell fractionation procedures used. However, as mentioned in chapter 1, overproduced MtgA was identified in a precipitate (at 200 000 × g) fraction of W3110 cells by Di Berardino *et al.* [Di Berardino *et al.*, 1996], which should correspond to total membrane fraction in Fig. 4.8. Di Berardino *et al.* used 30 μg of the precipitate fraction per track on a gel to be able to visualise the MtgA band with Coomassie staining. In Fig. 4.8, an amount of protein loaded per every track corresponds to about 12 μg of total cell protein, of which the precipitate fraction constitutes only a very small fraction. The faster-moving band which presumably corresponds to the P8-'ScrP chimaeric protein is found in the cytoplasmic fraction only.

In the case of pTAK11, the three overproduced bands present in total cell extract are also found in cytoplasmic and periplasmic fractions, but are absent from the total membrane fraction (Fig. 4.9). This confirms the theoretical prediction that ScrP should be transported into the periplasm. The presence of all the three bands in the *c* and *p* tracks suggests that none of them are MtgA (cf. pTAK8 results), and careful mobility comparisons support this view. Presumably MtgA is present in the total cell extract of the pTAK11, but too faint to be detected.

Figure 4.8. 10% tricine-SDS- PAGE of proteins from cells overexpressing ScrP and MtgA (from pTAK11) or MtgA and P8-'ScrP (from pTAK8), reverse-stained. *M*, molecular size marker; *t(–I)*, total cell extract, culture not induced with IPTG. Cells induced with IPTG: *t*, total cell extract; *c*, cytoplasmic fraction; *p*, periplasmic fraction; *tm*, total membrane fraction; *om*, outer membrane fraction.

The band patterns seen for ScrP overproduced by pTAK11 vary in mobility and intensity between two different gels (Fig. 4.7 and 4.9) as well as in intensity between total cell extract and periplasmic and cytoplasmic fractions (Fig. 4.9). The two different gel systems used might be partially responsible for the different band patterns, but could not explain their complexity. To investigate the possibility that the bands represent different stages of processing of ScrP protein, a pulse-chase labelling experiment was carried out.

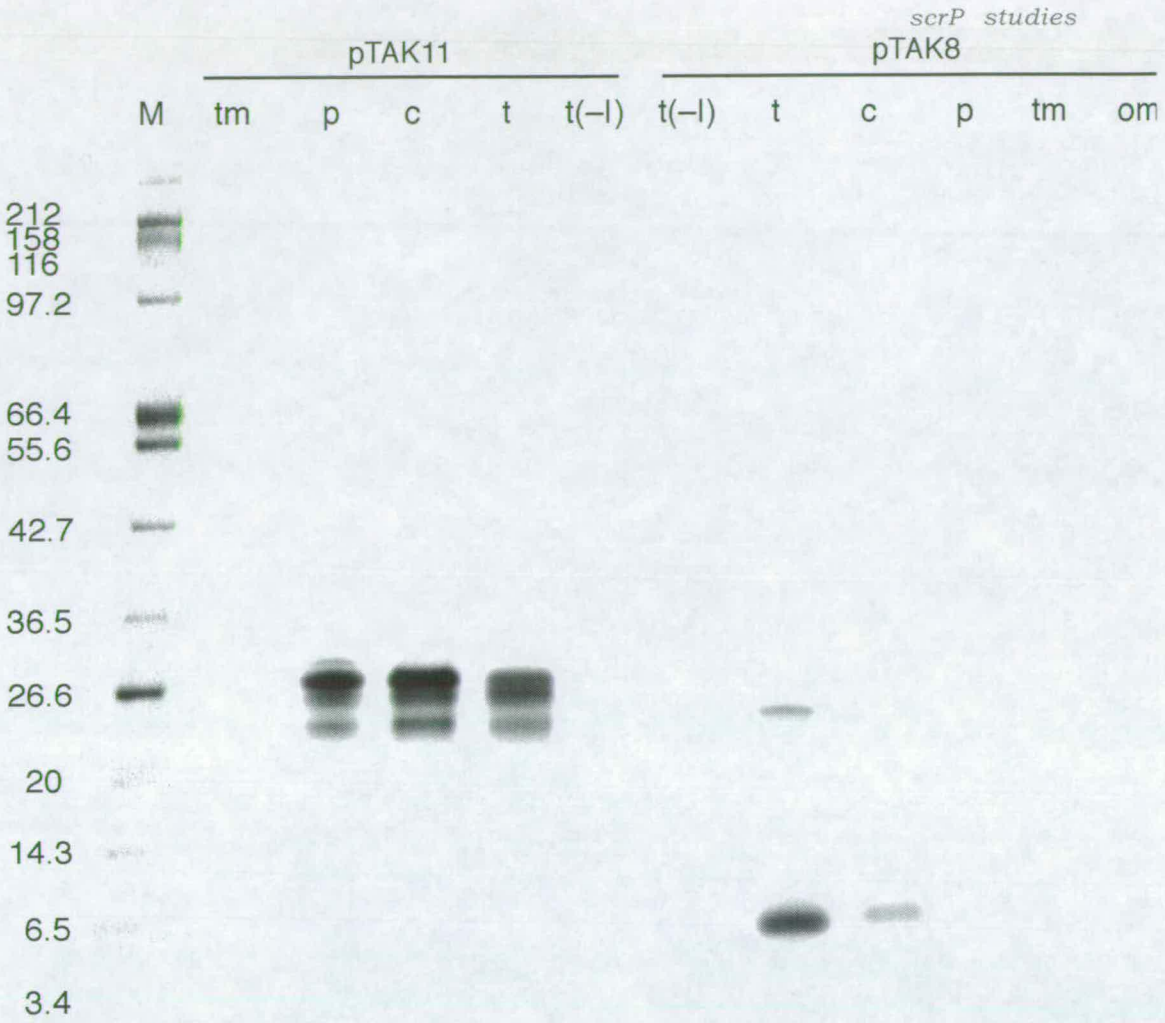


Figure 4.9. Autoradiograph of the gel shown in Fig. 4.8. The marker track is Coomassie-stained.

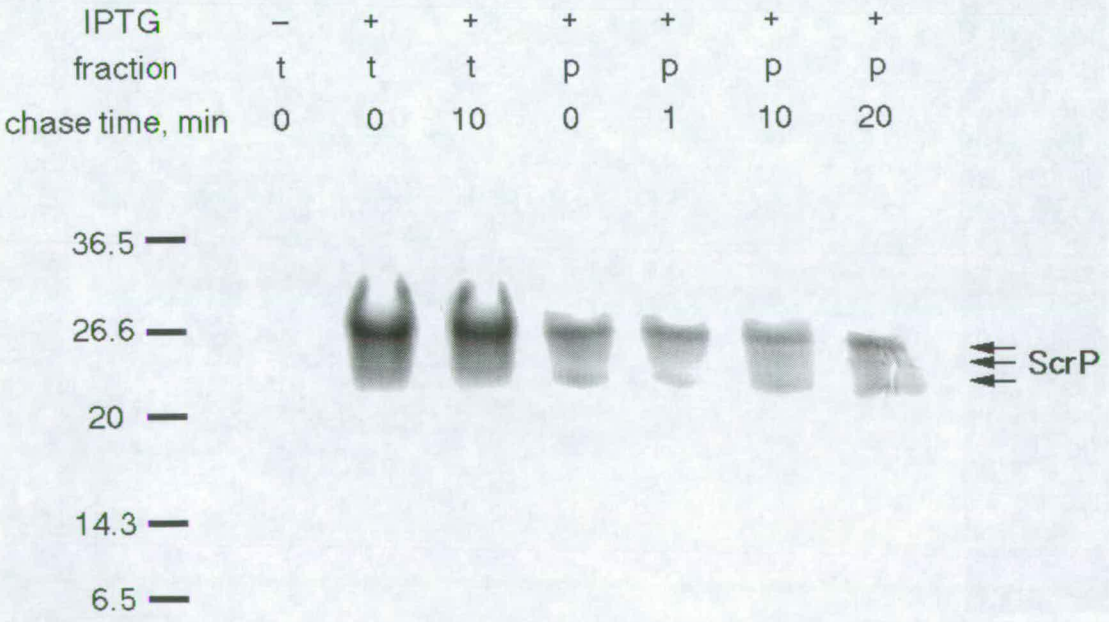


Figure 4.10. Fractionation of pulse-chase labelled proteins overproduced from pTAK11 (ScrP + MtgA) on 10% Glycine-SDS-PAGE gel. Cells were labelled according to a protocol described in Section 2.5.3 for 5 minutes, then cold methionine was added and culture incubation was continued for various times, as shown. IPTG+/–, with or without induction with IPTG; t, total cells; p, periplasmic fraction.

As seen in Fig. 4.10, the relative intensity of the ScrP bands did not change detectably during the chase of up to 20 minutes. This suggests that the bands are not forms of ScrP at different stages of processing. While it is possible that a different labelling protocol could produce a less negative result, the fact that the band patterns in the cytoplasmic and periplasmic fractions of pTAK11-containing cells are identical (Fig. 4.9) also argues against the idea that the different bands are due to cleavage of ScrP during transfer to the periplasm. The appearance of the minor bands may be an artefact of the overproduction of ScrP, and a consequent inability of the cells to cope with its incorrect folding, leading in turn to some instability.

4.3. Analysis of expression of the genes in the *arcB* operon.

To investigate the expression of the *scrP* and *mtgA* genes *in vivo*, whether it is dependent on that of *arcB*, and how *scrP* and *mtgA* are expressed under various growth conditions, it was decided to construct translational fusions of *arcB*, *scrP* and *mtgA* to *lacZ*, and then transfer these onto the *E. coli* chromosome, replacing the natural *arcB*, *scrP* or *mtgA* genes. This approach avoids artefacts due to abnormal localisation of the fusions, as well as ensuring a normal gene copy number.

4.3.1. Cloning of the *Eco*R I–*Bam*H I fragment of pACE1 into pUC18.

The chosen method of chromosomal gene replacement requires that the desired insertion, together with flanking homologous DNA, is present on a high copy number pUC-based plasmid. The *Eco*R I–*Bam*H I fragment of pACE1 was therefore cloned into pUC18 digested with the same enzymes. The structure of the resulting plasmid, pBRU12, was confirmed by restriction mapping with *Bam*H I; *Bst*B I; *Mlu* I and *Bgl* II; *Cla* I and *Bgl* II; and *Bst*B I and *Bgl* II.

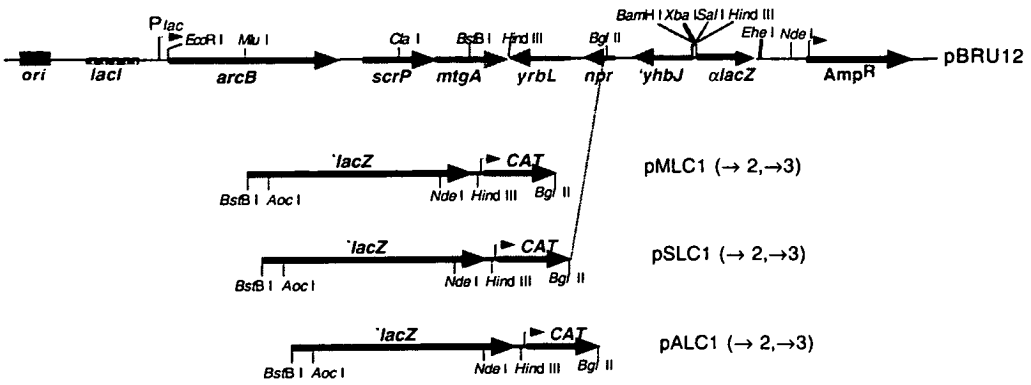


Figure 4.11. Scheme of construction of pALC3, pSLC3. pMLC3. Relevant restriction sites are shown. *Xba* I–*αlacZ–Ehe* I is present only in pBRU12, pALC1, pSLC1, and pMLC1.

4.3.2. *lacZ* fusion construction

Each fusion was constructed in such a way that all the *E. coli* DNA downstream from the point of fusion in *arcB*, *scrP*, or *mtgA*, and extending as far as the first *E. coli Bgl II* site beyond the end of *mtgA* was removed, and replaced with the full-length *lacZ* gene, followed by a CAT cassette encoding chloramphenicol resistance gene (including its own promoter) (Fig. 4.12).

The resulting ArcB'–'LacZ fusion protein would include the first 379 amino acids of ArcB ending in the middle of the H1 domain (amino acids 269–515) and thus containing His-292, but lacking the G-rich and N-domains necessary for autophosphorylation (see chapter 1). Similarly the ScrP'–'LacZ and MtgA'–'LacZ fusion proteins would have about half of ScrP or MtgA respectively, fused to LacZ.

First the *lacZ*–CAT cassette was amplified by PCR. A DNA template carrying the full length *lacZ* followed by CAT cassette, and primer 3'Cml corresponding to the downstream end of the *lacZ*–CAT cassette, were kindly provided by Nicola Holden. Primers for creating suitable restriction sites at the beginning of the *lacZ* were designed for each fusion: ASCI-LACZ to allow ArcB'–LacZ fusion,

Bscrp for ScrP'–LacZ, and Bmtga for MtgA'–LacZ fusions (listed in chapter 2.1).

PCR was conducted as described in Section 2.4.13, using Vent DNA polymerase in order to minimise PCR generated mutations, with annealing temperature of 55°C and extension time of 3 min, appropriate for the length of the amplified DNA. The 4.0 kb of PCR products were gel-purified, then digested with *Asc* I and *Bgl* II (for the *arcB* fusion construction), or *Bst*B I and *Bgl* II (for the *scrP* and *mtgA* fusion constructions). *Asc* I and *Mlu* I produce compatible cohesive ends, as do *Cla* I and *Bst*B I. The fragments were accordingly ligated to pBRU12 digested with *Mlu* I and *Bgl* II (for the *arcB* fusion construction), *Cla* I and *Bgl* II (for the *scrP* fusion construction) or *Bst*B I and *Bgl* II (for the *mtgA* fusion construction), respectively. However, attempts to clone these PCR products directly into pBRU12 failed, and the whole PCR products were first subcloned into pBR322 digested with *Eco*R I (for the *arcB* fusion construction), or *Nru* I (for the *scrP* and *mtgA* fusion constructions). They were digested with the enzymes as above (*Bgl* II with *Asc* I or *Bst*B I), 4.0 kb fragments were purified and then cloned into pBRU12.

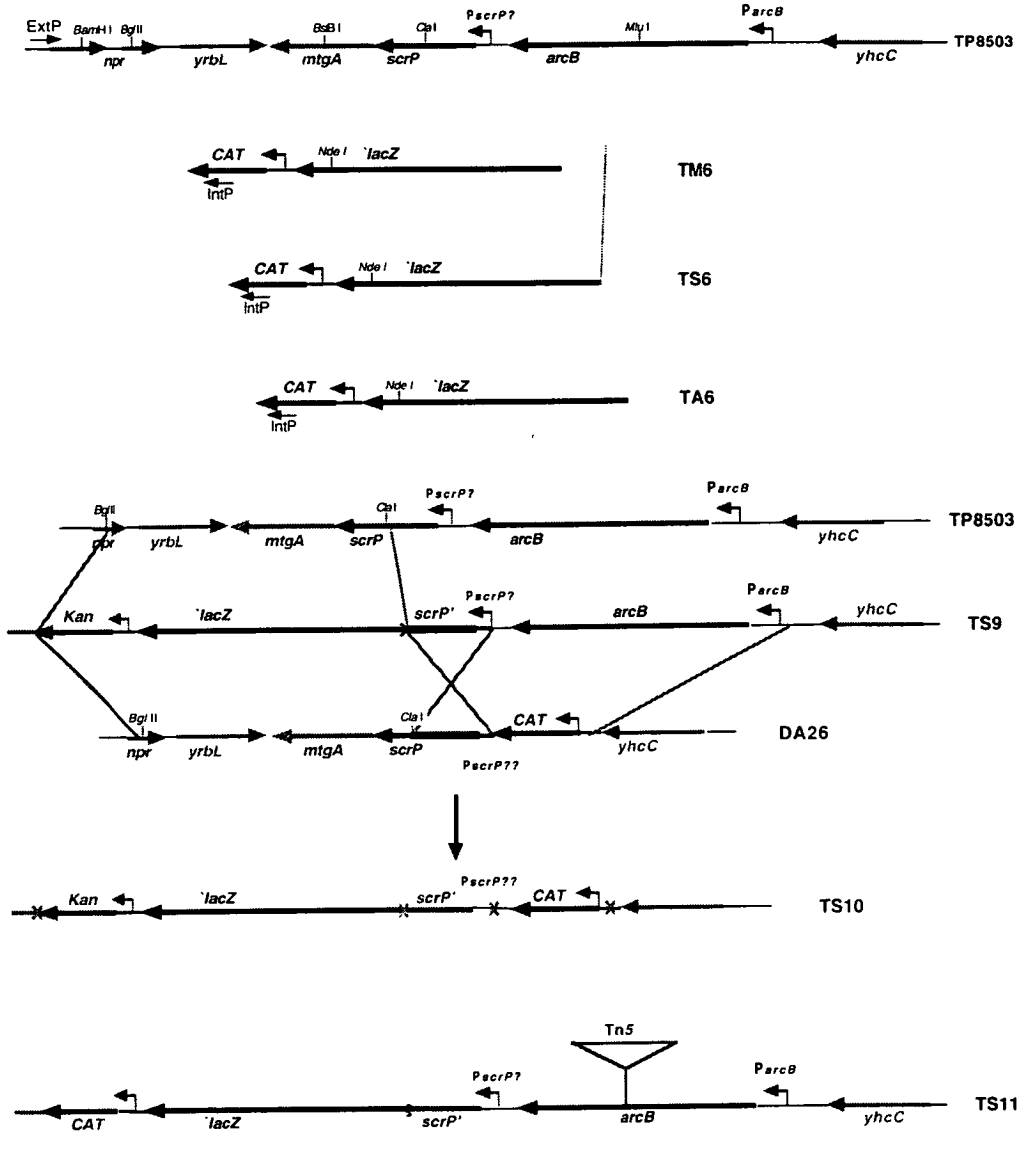


Figure 4.12. The scheme of chromosomal *lacZ* fusions constructions. Relevant genes and restriction sites are shown. The crossover between homologous sequences of the chromosomes of TS9 and DA26 is about 520 bp. The orientation of genes in Tn*5* is not known.

After transformation into the *E. coli* strain DH5α, Cm^R and Amp^R colonies growing at 37°C and turning blue on LB–plates containing X-gal were chosen for analysis.

The plasmids were named pALC1, pSLC1, and pMLC1 — for *arcB*, *scrP* and *mtgA* fusions to *lacZ*–CAT, respectively. Their structures were confirmed by restriction analysis with *Bgl* II, *Eco*R I (Fig. 4.11).

pUC18 contains the sequence coding for the α-peptide of β–galactosidase. To ensure stability of the fusion constructions in *recA*$^+$ cells it was necessary to remove this sequence. To achieve this, pALC1, pSLC1 and pMLC1 were digested with *Ehe* I and *Xba* I (Fig. 4.11). The recessed 3' termini were then filled in by treatment with Sequenase® Version 2.0 T7 DNA polymerase, and re-ligated. After ligation the mixture was digested with *Sal* I endonuclease, to reduce recovery of the parental plasmid. The products were then transformed into DH5α and the transformants were plated at 30°C. The plasmid products were checked for the absence of *Xba* I, *Ehe* I and *Sal* I sites and the second *Hin*d III site. The resulting plasmids were 186 bp shorter than their parents and were named pALC2, pSLC2, and pMLC2.

4.3.3. Replacement of the chromosomal copies of *arcB*, *scrP* and *mtgA* with their *lacZ* fusions.

The *arcB* region on the chromosome of *E. coli* strain TP8503 was partially replaced with the above *lacZ* fusion constructions according to the protocol described in section 2.3.6. TP8503 cells were transformed with pALC2, pSLC2 and pMLC2 and subjected to Amp^R and Cm^R selection at 30°C. The purified transformants were then grown in L-broth in the presence of chloramphenicol only, and P1 lysates were prepared on them. These lysates were used to transduce MM20 cells, which are *polA* and therefore cannot replicate ColE1 derived plasmids, to Cm^R. The resulting Cm^R Amp^S colonies were purified and used to produce second P1 lysates, which in turn were used to transduce TP8503 cells to Cm^R. However, none of the resulting Cm^R Amp^S colonies turned blue on

LB–plates containing X-gal, even though DH5α cells transformed with the corresponding plasmids had developed a blue phenotype. In principle this might be due to a copy number effect. On the other hand, the chromosome of DH5α cells contains the downstream portion of *lacZ* and produces the ω-peptide of β–galactosidase, whereas TP8503 is a *lacZ*-null strain. Therefore, it is possible that the PCR-generated *lacZ-ω* portions in pALC2, pSLC2 and pMLC2 contained errors producing a Lac^- phenotype, although their *lacZ-α* portions must be Lac^+ because the DH5α cells transformed with these plasmids had been screened for the blue colour on LB–plates containing X-gal.

To correct the *lacZ* sequence of pSLC2, its *Aoc* I–*Nde* I portion of *lacZ* was replaced with that from pML_{HM}::*lacZ*, a plasmid with the wild-type *lacZ* sequence from pTAC3953 [Tove Altung, *pers. comm.* to R. Hayward]. The sequence of the downstream end of *lacZ* beyond the *Nde* I restriction site in pSLC2 had been confirmed beforehand by sequence analysis of pSLC2 using primer 839F, which is directed upstream from within the beginning of the CAT gene. DH5α was transformed to Cm^R with the pSLC2/$lacZ^+$ ligation products. Transformants which turned blue on LB–plates containing X-gal were grown at 30°C, and had plasmid DNAs were isolated. TP8503 cells were then transformed with these plasmids and grown at 30°C. Blue TP8503 transformants carried pSLC3, the pSLC2 analogue with the wild-type *lacZ* sequence.

Next, pSLC3 was digested with *Aoc* I and *Bam*H I, and its large fragment was purified and ligated with the small *Aoc* I–*Bam*H I fragments of pALC1 and pMLC1. The ligation mixtures were enriched for the desired products by digestion with *Sal* I. Both the resulting plasmids should carry wild-type *lacZ*. They were named pALC3 and pMLC3, respectively (Fig. 4.11).

pALC3, pSLC3 and pMLC3 were transformed into *E. coli* TP8503 cells, and blue colonies were selected on LB–plates containing X-gal, Cm and Amp at 30°C. These were purified and grown in L-broth

at 30°C in the presence of chloramphenicol only, and used for the preparation of P1 lysates.

MM20 cells were transduced with the resulting lysates, plated on L–agar containing chloramphenicol only, and grown at 37°C. Colonies sensitive to ampicillin were then purified (TA5, TS5, and TM5) and used to produce fresh P1 lysates, which were used to transduce TP8503 cells. Transductant colonies which had a Cm^R Amp^S phenotype were selected at 37°C. The resulting strains were named TA6 (with the *arcB–lacZ* fusion), TS6 (*scrP–lacZ* fusion), and TM6 (*mtgA–lacZ* fusion) (Fig. 4.12).

In the presence of X-gal at 37°C, TA6 and TS6 colonies turn blue in their centres by the second day, whereas TM6 colonies do not give any noticeable blue colour.

4.3.4. Confirmation of the structure of TA6, TS6 and TM6.

The correct insertion of the *lacZ* fusion – CAT cassettes into the chromosome of TP8503 was confirmed by PCR analysis and by Southern hybridisation – the latter only for the TP8503, TA6 and TS6 strains.

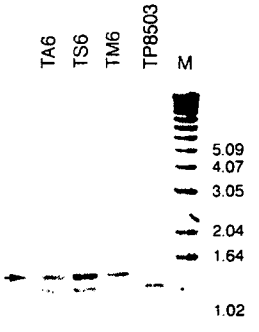


Figure 4.13. CAT is present on TA6, TS6 and TM6 chromosomes in the expected position. PCR analysis of single colonies of each of the strains using primers IntP and ExtP (Fig. 4.11). TP8503 (parental strain, CAT^-) is a negative control.

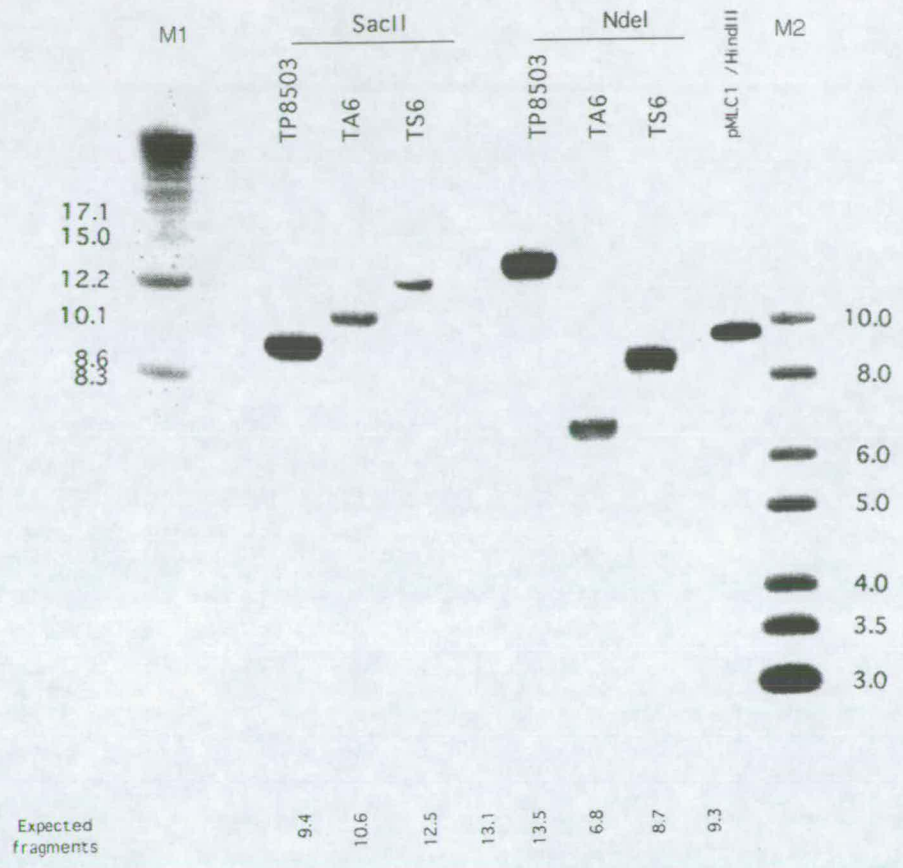


Figure 4.14. Southern blot analysis of chromosomal DNA of strains TP8503. TA6 and TS6 digested with indicated enzymes. The probe was [^{32}P]-labelled mixture of 3.96 kb *Eco*R I–*Hin*d III fragment of pTAK1 (100 ng) and λ DNA (10 ng) (Boehringer Mannheim) (see section 2.4.9). pMLC1 digested with *Hin*d III presented as a positive control. The markers are *M1* – Gibco High Molecular Weight Marker, *M2* – MBI Fermentas's 1 kb ladder. Fragment sizes are given in kb. Expected fragment sizes are shown at the bottom.

The PCR was performed on resuspended colonies of TA6, TS6 and TM6 cells, using primers IntP and ExtP to amplify a fragment of 1.45 kb, between the downstream end of the CAT gene and the segment of chromosomal DNA lying beyond the *Bam*H I site of the *Eco*R I–*Bam*H I fragment used to construct pBRU12 (Fig. 4.12). The expected product of this reaction can only be amplified if the CAT gene is inserted in the correct location on the chromosome, because the IntP primer is complementary to the plasmid borne CAT gene (not present on the TP8503 chromosome), and the ExtP

primer is complementary to a piece of chromosomal DNA not present on the plasmids (Fig. 4.11). All three fusion strands (TA6, TS6 and TM6) yielded this PCR product, whereas TP8503 did not (Fig. 4.13).

The analysis by Southern hybridisation was performed using chromosomal DNA purified according to the protocol described in section 2.4.10. The DNA was digested with the *Sac* II or *Nde* I restriction endonucleases, fractionated on a 0.7% agarose gel, Southern-blotted and probed with the gel-purified 3.96 kb *Eco*R I–*Hin*d III fragment of pTAK1 which includes the *arcB*, *scrP* and *mtgA* genes, [^{32}P]-labelled as described in chapter 2 (2.4.9, 2.4.12).

The expected size of the fragment hybridising to the probe in the case of the *Sac* II digests is 9.4 kb for TP8503, 10.6 kb for TA6, and 12.5 kb for TS6 (Fig. 1.2.1). In the case of the *Nde* I digests, the expected size is 13.5 kb for TP8503, 6.8 kb for TA6, and 8.7 kb for TS6.

The sizes of the fragments responding to the probe in the hybridisation experiment clearly correspond to those expected, thus confirming that the insertions of the *lacZ*–CAT cassettes have occurred in the correct loci (Fig. 4.14).

4.3.5. Sensitivity to dyes.

One phenotype produced by *arcA* and *arcB* mutations is sensitivity to certain redox dyes, such as toluidine blue and methylene blue. The mutants give tiny or no colonies on plates containing these dyes [Ishige *et al.*, 1994]. TA6 should be ArcB$^-$ due to insertion of the *lacZ* cassette in *arcB*, whereas TS6 and TM6 should be ArcB$^+$. These predictions were tested as follows.

MM20 and DA26 are strains which carry *arcB* mutations [Smillie, 1994; Ishige *et al.*, 1994]. TA6, as well as MM20 and DA26, proved sensitive to toluidine blue and methylene blue when grown on T-agar in the presence of dye. Moreover, when MM20 (*arcB*) was

transduced with P1 lysates grown on TS6 or TM6 strains and selected for Cm^R, all of the resulting transductants proved resistant to toluidine blue and methylene blue. These results provide further evidence for the correct insertion of the *lacZ*–CAT cassette into the TA6, TS6 and TM6 chromosomes.

4.3.6. Complementation tests.

Mutations in *arcB* can be complemented by providing $arcB^+$ on a plasmid [Elding, 1995]. Tests of this sort were therefore carried out, as summarised in Table 4.2.

Table 4.2. Colony size of strains on dye agars (toluidine blue and methylene blue) at 37°C. TP8503 is *arcB*+ control strain and, and pJJW11 is the *arcB*-null control (vector) plasmid. —: not tested.

Strain	TP8503	TS6	DA26	TA6
Plasmid —	normal	normal	tiny	tiny
pJJW11	—	—	tiny	tiny
pACE1	—	—	normal	tiny
pTAK1	—	—	normal	tiny
pFMT1	—	—	normal	tiny

It can be seen that DA26 cells (*ΔarcB::*CAT) carrying pTAK1, pACE1 or pFMT1 all grow well on toluidine blue and methylene blue at 37°C. This indicates that *ΔarcB* can be complemented in *trans* by all three $arcB^+$ plasmids, including pFMT1. This further confirms that pFMT1 encodes a fully functional ArcB, in contrast to earlier concern (mentioned in chapter 1) that pFMT1 might carry a truncated *arcB* which could not produce a functional protein [Smillie, 1994].

In contrast, TA6 cells (*arcB'–'lacZ,* CAT) transformed with either pJJW11, pTAK1, pACE1 or pFMT1 produce tiny colonies on dye agars. This indicates that the *arcB* mutation in TA6 can not be complemented in trans.

TA6 cells grow more slowly even on dye-free agar than those of TS6 and TM6, which carry *scrP–lacZ* and *mtgA–lacZ* fusions, or their parent TP8503. This suggests that the fusion protein ArcB-LacZ may be toxic to cells. To test if this is the case, strains were transformed with plasmids encoding *arcB* + *scrP* + *mtgA*, (pACE1, pTAK1 and pBRU12), or plasmids carrying the *arcB–lacZ* or *scrP–lacZ* fusions together with the CAT gene (pALC3 and pSLC3) and grown on the dye (toluidine blue) agar at 30°C. The results are presented in Table 4.3.

The reason for doing these dye-sensitivity tests at the lower temperature was the previously observed deleterious effect of the pUC18 derivatives pBRU12, pALC3, pSLC3 and pMLC3 on colony growth at 37°C, presumably due to the very high copy number of these plasmids and consequent overexpression of the membrane protein LacZ'(8aa)–ArcB or its 'LacZ fusion (ArcB produced from pBRU12 and its derivatives has 8 extra amino acids at its *N*-terminus, due to fusion to the beginning of LacZ). Similar observations of the toxicity of ArcB and even of ArcB lacking its transmembrane domains, when overexpressed from very high copy number plasmids (derivatives of pBluescript), were reported earlier [Iuchi and Lin, 1992b; Iuchi and Lin, 1992a]. In contrast, all the plasmids shown in Table 4.3. were derived from pBR322, and their effect on growth rate of cells at 30 or 37°C is negligible.

Table 4.3. Colony size of strains on dye agar (toluidine blue) at 30°C. pACE1 and pTAK1 are derivatives of pBR322. pBRU12, pALC3 and pSLC3 are derivatives of pUC18. pACE1, pTAK1, pBRU12 and pSLC3 contain full-length *arcB*. TP8503 and DH5α are *arcB*$^+$. —: not tested.

Strain	TP8503	TA6	DH5α
Plasmid			
—	normal	tiny	normal
pACE1	—	tiny	—
pTAK1	—	—	normal
pBRU12	—	tiny	tiny
pALC3	tiny	—	tiny
pSLC3	—	—	tiny

All the plasmids with very high copy numbers, including pBRU12 itself, show toxicity in all the strains analysed. Therefore, it is impossible to test the LacZ'(8aa)–ArcB'–'LacZ fusion protein itself for particular toxicity using the set of plasmids we currently possess. It might be useful to transfer the DNA encoding ArcB'–'LacZ fusion to a plasmid with the copy number of pBR322 or even lower, but this has not yet been done. The reason for the failure of $ArcB^+$ plasmids to complement the *arcB'–'lacZ* single copy fusion gene to Dye^R in strain TA6 remains unknown.

4.3.7. The expression levels of the *arcB–lacZ*, *scrP–lacZ* and *mtgA–lacZ* fusions under various growth conditions.

β-galactosidase activity measurements.

ArcB, as part of a two-component regulatory system, is involved in the aerobic/anaerobic regulation of a large number of genes. It therefore seemed interesting to study the expression of all three genes in the putative *arcB* operon at different stages of growth under aerobic and anaerobic conditions. The strains carrying the *lacZ* fusions were grown aerobically (with vigorous shaking) or anaerobically (under strict anaerobiosis, as described in section 2.2.4) in L-broth supplemented with 40 mM glucose, in the presence of the appropriate antibiotics (section 2.2.4). Samples were taken at different times during growth, and the β-galactosidase activity measured as described in section 2.5.11. Raw data for one strain and growth conditions are exemplified by Fig. 4.15, and the overall results are presented in Fig. 4.16.

Under aerobic conditions, TA6 cells produce a β-galactosidase activity of about 11 Miller units during exponential growth. Upon entry into stationary phase, under the same conditions, which rises about 2-fold soon after entry into stationary phase (Fig. 4.15, 4.16 *a*, and Table 4.4).

TS6 cells, grown under the same conditions, produce a β-galactosidase activity of about 9 Miller units during exponential growth, which rises 1.55-fold soon after entry into the stationary phase (Fig. 4.16 *a*, table 4.4).

As mentioned before, Iuchi and Lin [Iuchi and Lin, 1992a] reported that the expression of *arcB* does not differ significantly between aerobic and anaerobic conditions. Our data show that it is about 1.4 times higher under anaerobic than under aerobic conditions in the exponential growth phase, but is practically the same in the stationary phase under the both conditions (Fig. 4.16 *b*). This might suggest that the latter increase under aerobic conditions could be, at least partly, due to oxygen starvation at high cell densities. Although these changes in activity are rather small, they appear to be statistically significant (Table 4.4).

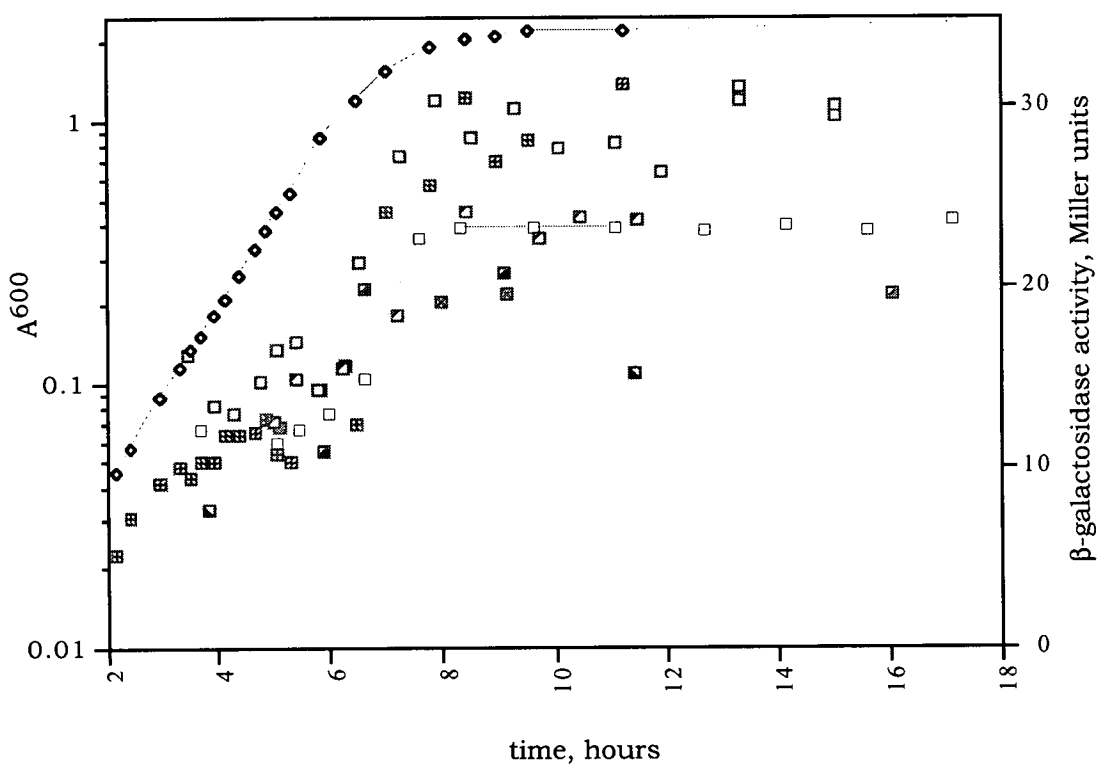


Figure 4.15. Growth curve and specific β-galactosidase activity of TA6 cells grown on glucose-enriched L-broth under aerobic conditions. A^{600} (left axis, diamonds) and specific β-galactosidase activity (right axis, squares) are plotted against time. Different types of square symbols represent the results of 7 separate experiments. The data of one of the experiments (whose results are close to the mean) are connected by a dotted line.

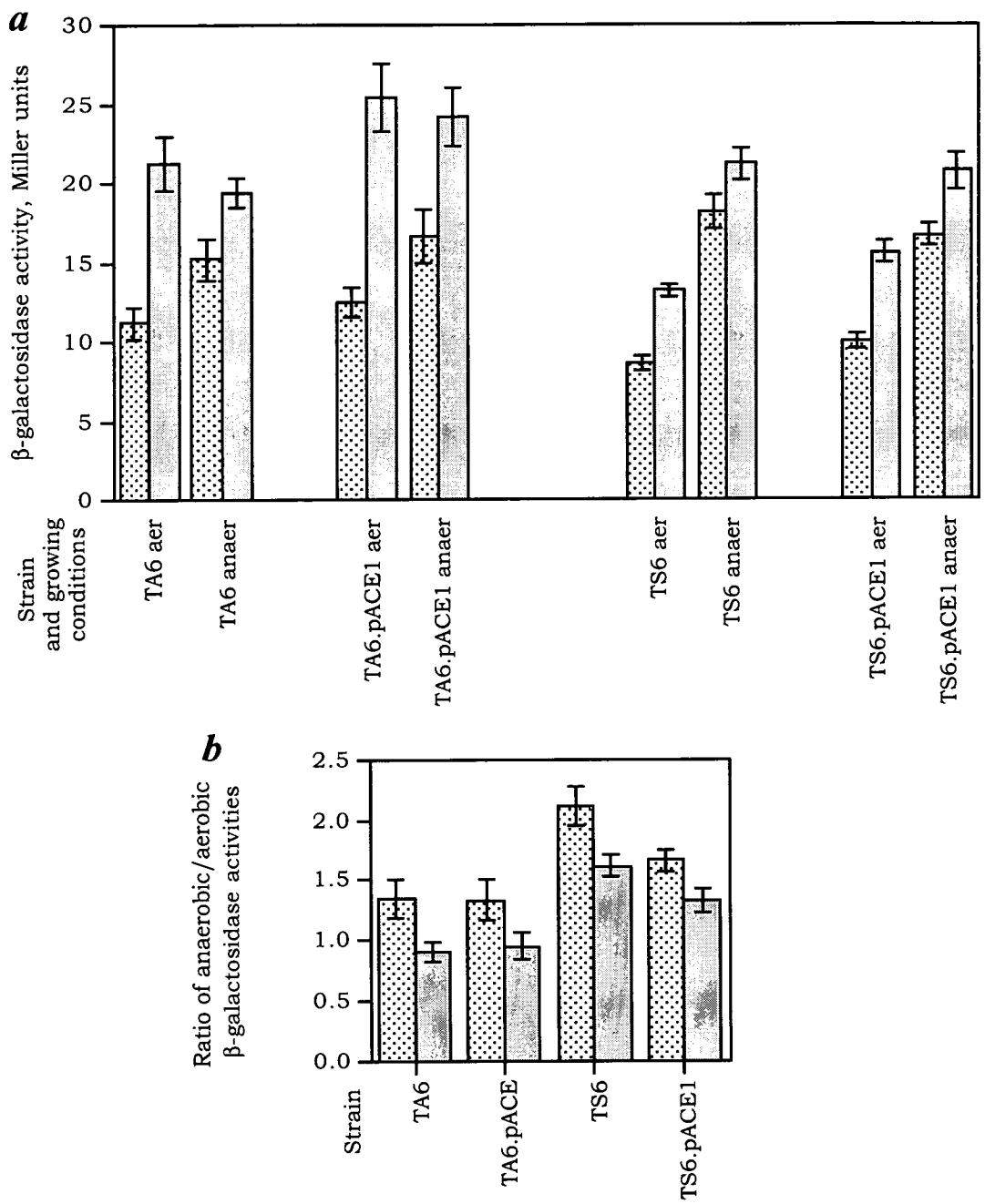


Figure 4.16. *a)* β-galactosidase activity of TA6 and TS6, with or without pACE1, grown under aerobic and anaerobic conditions, in Miller units. The results shown are the means and standard errors of the activities in exponential (▒) and stationary (□) phases, each derived from 6–11 experiments. A result of a single experiment is a mean of β-galactosidase activities (MU) of samples taken at A^{600} between 0.1 and 0.4 (for exponential phase); and a β-galactosidase activity of a sample taken 10 – 12 hours (aerobic growth) or 16–20 hours (anaerobic growth) after start of incubation.

b) Ratio of the β-galactosidase activity (MU) under anaerobic conditions to that under aerobic conditions, for different strains, calculated from the data in panel *a*.

Table 4.4. Data for Fig. 4.16, β-galactosidase activities (means and standard errors) of TA6 and TS6 under different conditions, in Miller units. The data are plotted in Fig. 4.16. Stationary/exponential phase ratios of β-galactosidase activities (the last column) are from the ratios of individual experiments.

Growth conditions	Growth phase	Exponential	Stationary	Stationary/ exponential ratio
Aerobic	TA6	11.2 ± 1.0	21.3 ± 1.7	1.99 ± 0.11
	TA6·pACE1	12.5 ± 0.9	25.4 ± 2.21	2.04 ± 0.17
	TS6	8.7 ± 0.4	13.2 ± 0.5	1.55 ± 0.04
	TS6·pACE1	10.1 ± 0.4	15.7 ± 0.8	1.56 ± 0.08
Anaerobic	TA6	15.2 ± 1.4	19.5 ± 1.0	1.29 ± 0.11
	TA6·pACE1	16.7 ± 1.7	24.3 ± 1.8	1.49 ± 0.12
	TS6	18.2 ± 1.1	21.3 ± 1.0	1.19 ± 0.08
	TS6·pACE1	16.7 ± 0.7	20.8 ± 1.2	1.24 ± 0.06
Anaerobic/ aerobic ratio	TA6	1.4 ± 0.2	0.9 ± 0.1	
	TA6·pACE1	1.3 ± 0.2	1.0 ± 0.1	
	TS6	2.1 ± 0.2	1.6 ± 0.1	
	TS6·pACE1	1.7 ± 0.1	1.3 ± 0.1	

In stationary phase TA6 cells produce similar amount of β-galactosidase under both aerobic and anaerobic conditions, whereas TS6 cells give specific activity higher under anaerobic than aerobic conditions (Fig. 4.16 *a*, table 4.4).

Under anaerobic condition (exponential growth) the expression of *scrP–lacZ* fusion is about 2.1 times higher than under aerobic conditions, against only 1.4 times increase for *arcB–lacZ*.

These observation provide a very tentative evidence for the following hypothesis. If the effects are transcriptional (this is not clear, since the protein fusions could also be affected by post-transcriptional regulation of *arcB* and *scrP*), then the *arcB* promoter is somewhat induced by oxygen starvation; *scrP* is cotranscribed

with *arcB*, but can also be served by its own independent promoter which is induced under anaerobic conditions.

The TM6 strain, which carries the *mtgA–lacZ* translational fusion on the chromosome, does not exhibit any detectable β-galactosidase activity. This could either be due to its very low expression, or to inactivation of β-galactosidase as a result of an abortive attempt by the cell to transport the fusion protein through the inner membrane into the periplasm – the normal locale of the MtgA protein [Di Berardino *et al.*, 1996]. The former is possible because even the best estimates of MtgA synthesis from T7 RNA polymerase-driven overexpression experiments suggest that there can be 7–10 times less MtgA synthesised than ScrP. The measured β-galactosidase activity of ScrP–LacZ fusion protein is 7–15 Miller units, therefore MtgA-LacZ fusion protein, provided it is active, would not be expected to be expressed at a detectable level.

Growth curves show that the doubling time for ArcB$^-$ strain TA6 is approximately 1.5 (1.49 ± 0.15) times higher than that for its isogenic ArcB$^+$ counterpart, TS6, at least under aerobic conditions. However, when both strains carry complementing pACE1 or pTAK1 with the *arcB–scrP–mtgA* served by a plasmid promoter the ratio of their doubling time becomes similar (1.02 ± 0.18). Thus the ArcB$^-$ phenotype of TA6 is complemented by these plasmids in terms of growth rate, making it all the more puzzling that its dye-sensitivity is not complemented (as mentioned in the previous section). Since the dye-sensitivity of strain DA26 is complemented, it seems likely that the nature of the ArcB–LacZ fusion protein produced in TA6 that underlies its persistent sensitivity. Perhaps the latter phenotype is more sensitive to the presence of defectively folded chimaeric ArcB protein in the cell, or its inappropriate localisation within the cell membrane. This can be especially the case if ArcB acts as a dimer when it senses the signal (in its sensor part) and passes it down to kinase domains.

The β-galactosidase expression by TA6 is scarcely affected by the presence or absence of pACE1, and the expression by TS6 shows

little more effect (Fig. 4.16 *b*). This indicates that the expression of *arcB* and *scrP* is ArcB-independent, and the expression of *scrP* is largely, if not completely, ScrP-independent.

Detection of fusion proteins measured by Western analysis.

To confirm that the β-galactosidase measurements truly reflect the level of transcription and translation of the fusion genes and are not affected by decay of their protein products, cell extracts were subjected to SDS–PAGE and Western analysis.

TA6, TS6 and TM6 cultures were sampled during aerobic growth. Samples were concentrated by centrifugation, resuspended in SDS-PAGE loading buffer, lysed and fractionated by 7% SDS–PAGE. The gel was then blotted onto a nitrocellulose membrane and probed with anti-β-galactosidase antibodies, as described in section 2.5.2. The results are shown in Fig. 4.17 and 4.18.

There is a relatively high background, presumably due to the low level of expression of the fusion proteins and not the very impressive specificity of the antibody. Nevertheless it is clear that no extensive accumulation of β-galactosidase decay products has been detected in the TA6 or TS6 samples (Fig. 4.17).

In the exponential phase samples, the band corresponding to the ScrP–LacZ fusion protein shown on Fig. 4.18 is at least 6-fold more intense than that of MtgA–LacZ, and ArcB–LacZ is about twice as intense as ScrP–LacZ (it is not possible to talk about amounts of protein expression here as there was no calibration for amounts β-galactosidase made in this experiment). These observations do not contradict to relative amounts of ScrP and MtgA produced in the T7 RNA polymerase-driven *in vivo*-overexpression system and with the β-galactosidase activity measurements discussed in the previous section. The difference between the ArcB–LacZ and ScrP–LacZ results could reflect partial termination of transcription between *arcB* and *scrP*, and/or a different efficiency of their ribosome-binding sites.

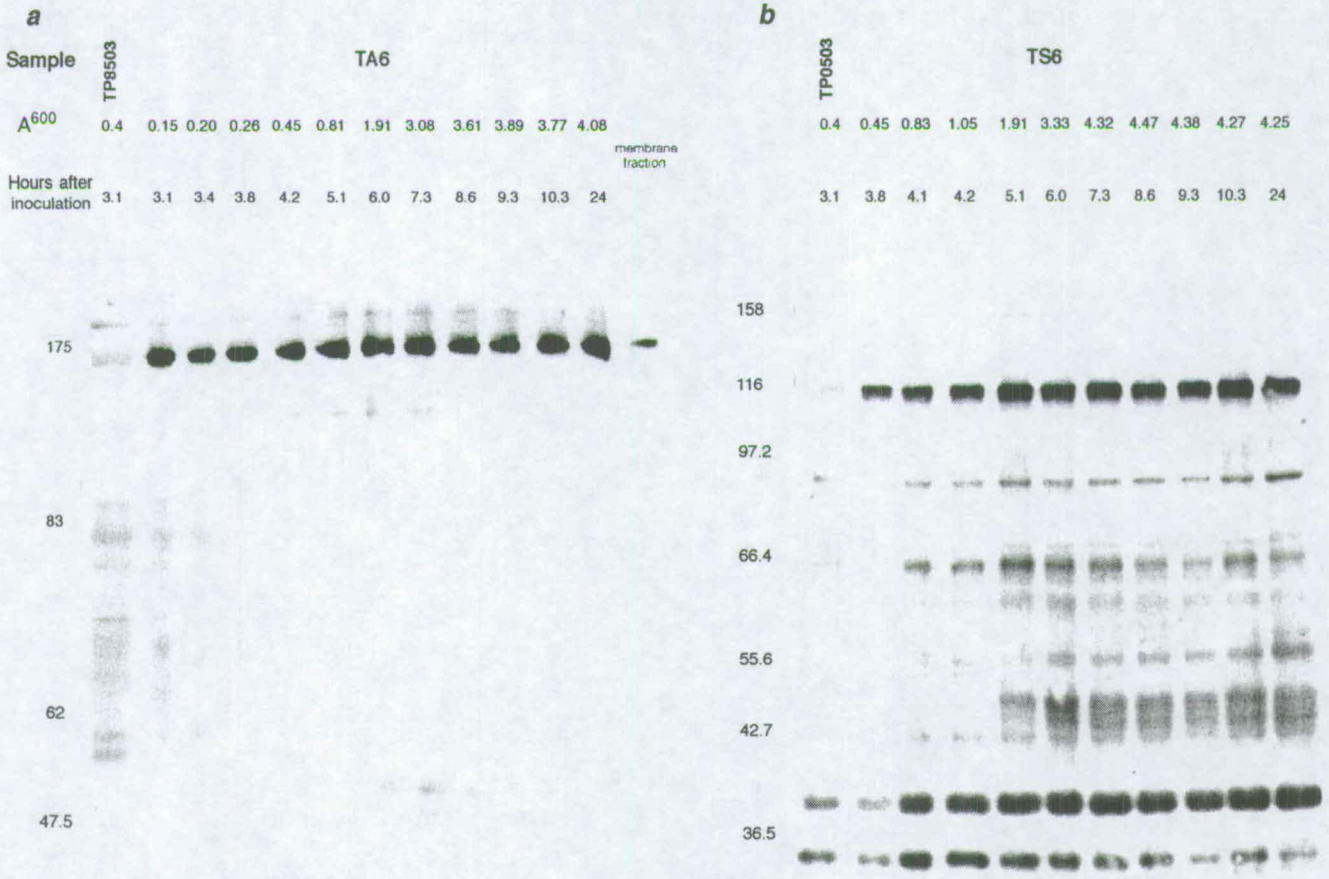


Figure 4.17. Western blots of SDS–PAGE fractionated extracts of cells sampled at different times and A^{600} from 37°C L-broth cultures of *a)* TA6 and *b)* TS6. The total protein inputs in each track were nearly equal. Molecular size markers and a TP8503 (LacZ-null) control track are shown on the left of each blot. A sample of total membrane fraction of TA6 taken at $A^{600}\approx3.9$ is also shown on the right of gel *a*.

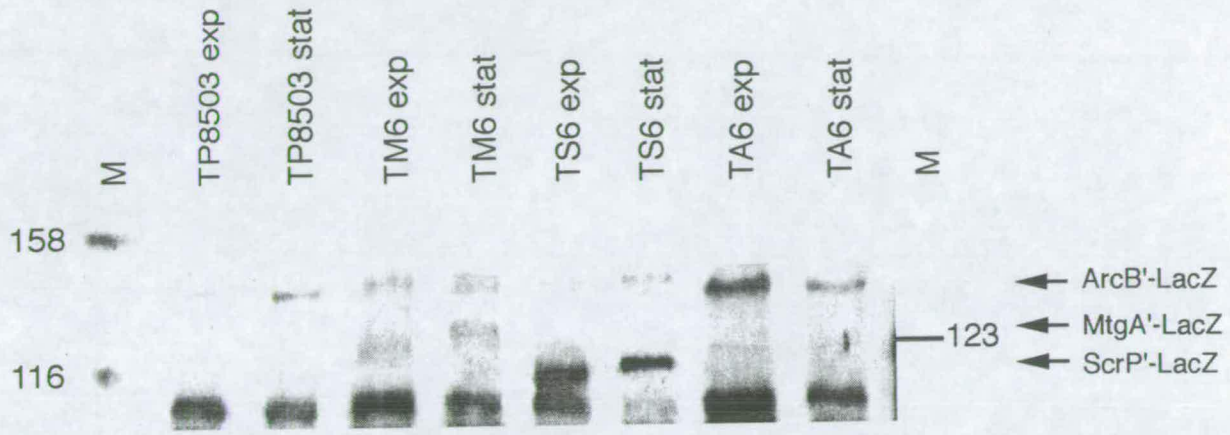


Figure 4.18. Western blot of SDS–PAGE-fractionated extracts of cells sampled in exponential (A^{600}=0.5) *(exp)* and stationary (A^{600}=3.5) *(stat)* phase (10 hours after inoculation). The total protein inputs in each track were nearly equal. *M*, positions of molecular weight markers. (Only the relevant, upper part of the gel is shown; the rest is heavily developed and uninterpretable, the specificity of the detection system was much lower than in the previous figure).

The ArcB–LacZ and MtgA–LacZ fusion proteins migrate on the SDS–PAGE gel at rates lower than their predicted sizes (162 and 132 kDa), but well above β-galactosidase. However, ScrP–LacZ (predicted molecular size 132 kDa) migrates at the same rate as β-galactosidase (116 kDa) (Fig. 4.18). If this is not a result of anomalous mobility (ScrP itself migrates close to the predicted rate), it could reflect rapid proteolytic separation of the ScrP and LacZ moieties of the fusion protein. Taking into account that ScrP has a leader peptide directing it to the periplasmic space, such an instability of the fusion protein seems to be fortunate in terms of accuracy of measurement of its expression using β-galactosidase activity assays, since partially transported hybrid protein might not have such activity.

The total membrane fractions of the TA6 and TS6 cells were isolated according to the protocol described in Section 2.5.8, and

were loaded, along with total cell extracts, onto 8% SDS–PAGE gels (TA6, Fig. 4.17 *a*; data for TS6 are not shown). The ArcB–LacZ fusion protein (of expected size) appeared in the insoluble (membrane) fraction, thus suggesting that ArcB is associated with the membrane, through the residues within the *N*-terminal 379 amino acids present in the fusion. This should be the inner membrane, as the fusion protein displays β-galactosidase activity. This is consistent with earlier data and the prediction of an *N*-terminal membrane anchor (chapter 1). The insoluble fraction of the TS6 cells does not respond to the antibody, which is not unexpected, provided the assumption about the instability of the ScrP–LacZ fusion protein is true.

4.3.8. *scrP* expression appears to be independent of *rpoS*.

Because both the *arcB–lacZ* and *scrP–lacZ* fusion strains show higher specific β-galactosidase activity in stationary phase, an attempt was made to check whether the expression of *scrP* was affected by σ^S, a sigma factor active in stationary phase and encoded by *rpoS*.

AF681 [Farewell *et al.*, 1996] was used as a source of an *rpoS*::Kan^R insertion mutation, which was transduced into TS6. The resulting Cm^R and Kan^R colonies were screened for *kat*$^-$ (RpoS$^-$) phenotype, as described in Section 2.5.12.

Judged by the β-galactosidase activity in exponential and stationary phases (Fig. 4.19) TS6 and TS6*rpoS* showed exactly the same pattern of ScrP–LacZ expression, suggesting that its expression is independent of the presence or absence of σ^S. TS6*rpoS* required longer incubation to achieve the same cell density as *rpoS*$^+$ strains. This can be due to smaller initial cell density of *rpoS*::Kan^R cells resulting from smaller cell density of their overnight cultures.

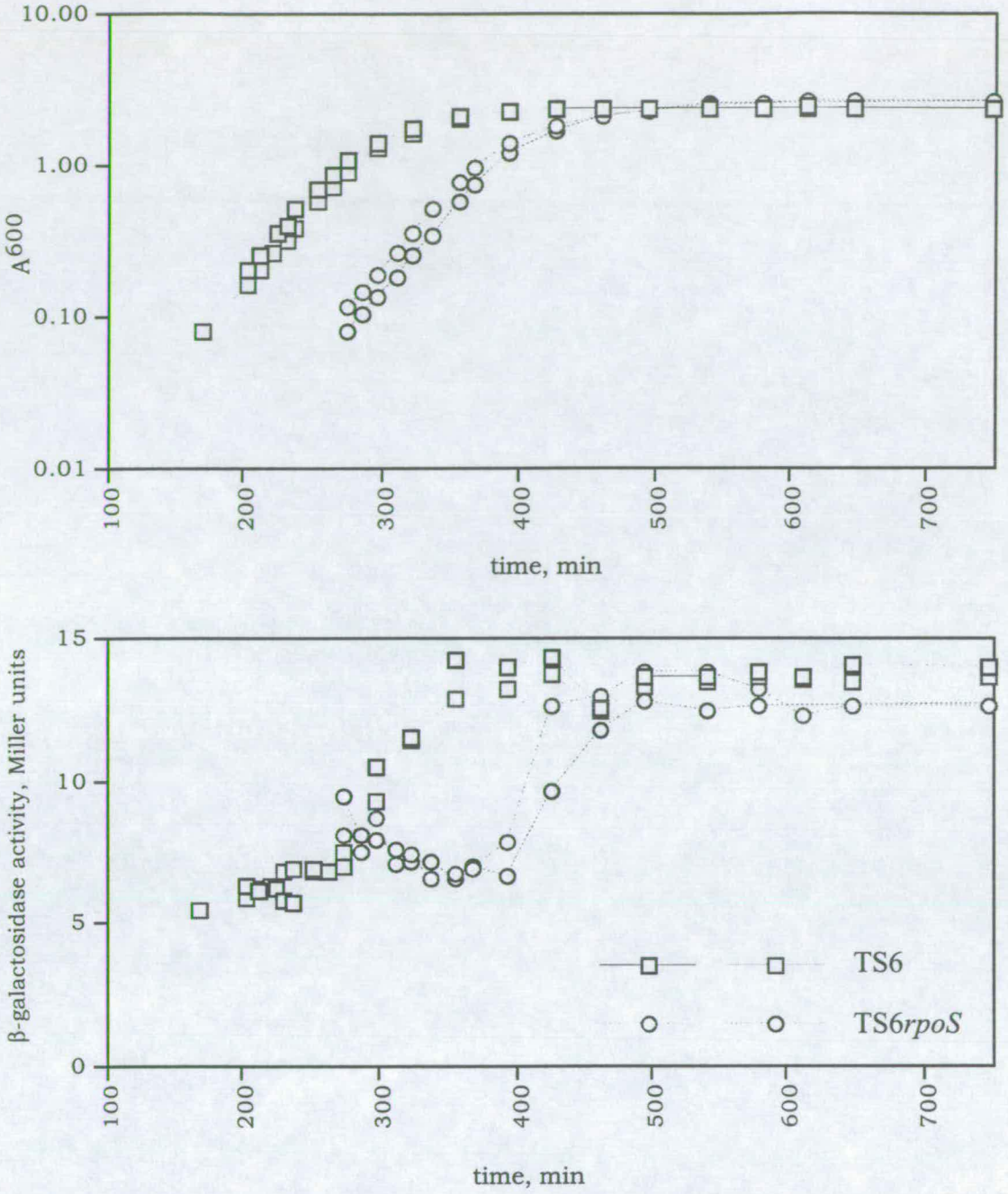


Figure 4.19. *a)* Growth curves and *b)* corresponding β-galactosidase activity measurements of TS6 (squares) and TS6*rpoS* (circles) at various times after dilution and growth in L-broth at 37°C under aerobic conditions. The results of two experiments are shown.

4.3.9. *arcB* and *scrP* expression is independent of *arcA*.

One study has suggested that *arcA* expression partly depends on the Arc system ([Compan and Touati, 1994], chapter 1). To test for

effects of Arc system on *arcB* and *scrP* expression, two *arcA* mutations were used. Strain ECL1601 was used as a source of the *arcA2* mutation linked to Tet^R (*zjj::*Tn*10*), and DS941 *arcA::*Tn*5* 2.3 was used as the source of a Tn*5* insertion within *arcA*. (Both strains were kindly provided by Dr. Sean Colloms). A P1 lysate grown on ECL1601 was used to transduce TP8503 to Tet^R. Tet^R dye-sensitive (*arcA2 zjj::*Tn*10*) and dye-resistant (*zjj::*Tn*10*) transductants were then purified and transduced to Cm^R with a P1 lysate grown on strain TA5. The resulting, purified dye-sensitive strains, TA8 and TA10, have the genotypes TP8503 *arcA2 zjj::*Tn*10 arcB'–'lacZ* CAT, and TP8503 *zjj::*Tn*10 arcB'–'lacZ* CAT, respectively.

The P1 lysate grown on TP8503 *arcA2 zjj::*Tn*10* cells was also used to transduce TS6, to produce the dye-sensitive strain TS12 (TP8503 *arcA2 zjj::*Tn*10 scrP'–'lacZ* CAT $arcB^+$) and dye-resistant TS13 (TP8503 $arcA^+$ *zjj::*Tn*10 scrP'–'lacZ* CAT $arcB^+$).

A P1 lysate grown on DS941 *arcA::*Tn*5* 2.3 was used to transduce TA6 and TS6 cells to Kan^R. The resulting strains, selected for Kan^R Cm^R and screened for dye-sensitivity, were purified and named TA9 (TP8503 *arcA::*Tn*5* 2.3 *arcB'–'lacZ* CAT) and TS14 (TP8503 *arcA::*Tn*5* 2.3 *scrP'–'lacZ* CAT), respectively.

The expression of the *arcB'–'lacZ* fusion is essentially the same in $arcA^+$ *arcB* and *arcA arcB* backgrounds (Table 4.5 and Fig. 4.16), at least under studied aerobic conditions. TA6 (*arcB'–'lacZ* $arcA^+$), TA6 supplemented with $arcB^+$ expressed from a plasmid, TA8 (*arcB'–'lacZ arcA2*) and TA6 *arcA::*Tn*5* 2.3 all show very similar β–galactosidase activities under aerobic conditions, in both the exponential and stationary phases. This indicates that *arcB* expression is independent of ArcA regulator. However, this interpretation is subject to a caveat, because the dye-sensitive phenotype of TA6 is not complemented by pACE1 or pTAK1. This is indicative of some impairment of Arc function in this strain, even though normal growth rate is restored by addition of those plasmids. Therefore all measurements of the expression of the *arcB'–'lacZ* fusion in this work, including those in Table 4.5, have been carried out with the

Arc system inactive (in plasmid-free TA6 –ArcB⁻), or impaired (in TA6 carrying *arcB*⁺ plasmids). This might conceivably conceal any requirement for ArcA.

Table 4.5. Results of β-galactosidase activity measurements (Miller units) of strains grown in L-broth at 37°C under aerobic conditions unless otherwise indicated. Cells were sampled at A^{600} of approximately 0.3 (exponential phase) or about 11 hours after inoculation (stationary phase). *: the mean and range of two experiments, others are result of a single experiment.

Strain	Exponential phase	Stationary phase
TA6 (*arcA*⁺)	11.2 ± 1.0*	21.3 ± 1.7*
TA9 (TA6 *arcA*::Tn*5* 2.3)	10.0 ± 2.2*	16.8 ± 2.7*
TA10 (TA6 *zjj*::Tn*10*)	14.4	18.0
TA8 (TA10 *arcA2*)	13.3	17.2
TS6	8.7 ± 0.4*	13.2 ± 0.5*
TS14 (TS6 *arcA*::Tn*5* 2.3)	6.8 ± 0.2*	9.9 ± 1.4*
TS13 (TS6 *zjj*::Tn*10*)	7.0	14.9
TS12 (TS13 *arcA2*)	9.0	17.0
TS10 (*scrP'–'lacZ*, Kan^R *ΔarcB*::CAT)	2.0 ± 1.0*	2.3 ± 0.2*
TS10, anaerobic growth	3.2	1.5
TS11 (TS6 *arcB*::Tn*5* 5.4)	3.6 ± 0.6*	5.4 ± 0.3*
TS11, anaerobic growth	7.7	5.7

No such caveat applies to the conclusion that *scrP* expression is clearly independent on ArcA. The expression of the *scrP'–'lacZ* fusion is independent of this regulator and, therefore, on Arc system, at least under aerobic conditions, because it is essentially the same in *arcB*⁺ *arcA*⁺ and *arcB*⁺ *arcA* strains (compare TS6 with TS14, and TS13 with TS12, in table 4.5).

4.3.10. Expression of the *scrP'–'lacZ* fusion in an *arcB* background.

The strain DA26 which carries *ΔarcB::*CAT has been used to make an *arcB*-null strain carrying a chromosomal *scrP'–'lacZ* fusion. DA26 has a complete deletion of *arcB*, including the likely *arcB* promoter, and of an undetermined amount of intergenic DNA upstream and downstream from *arcB* [Ishige *et al.*, 1994]. Before constructing the required strain it was necessary to find out if the length of DNA surviving on DA26 chromosome, between the CAT insertion and the *Cla* I site in the *scrP* gene (the site of the *lacZ* fusion insertion in TS6), would be sufficient for chromosomal recombination. This was done by PCR analysis using of several sets of primers (Fig. 4.20, 4.21).

CATout2 is a primer complementary to the beginning, and IntP to the end, of the CAT gene. Both are directed away from the CAT gene. 083V is a primer which starts at bp 378 of the coding sequence of *scrP*, and is directed upstream from it.

The results of the PCR reactions show that amplification of DA26 (and not the control TS6) samples with the IntP/083V pair of primers gives a specific product of approximately 0.52 kb, while PCR with the CATout2/083V pair does not (Fig. 4.21 *a*). This indicates that the CAT gene on the DA26 chromosome is oriented in the same direction as *arcB* and *scrP*, and that the distance between the downstream end of the CAT gene and the *Cla* I site in *scrP* gene (which is 470 bp downstream from the start of *scrP*) is close to 0.52 kb. This should include about 50, but no more than 100 bp of the DNA upstream of the *scrP* gene.

As seen from Fig. 4.21 *c*, PCRs with primer 997M do not produce products when DA26 sample is used as a substrate, whereas use of primer 834T leads to products of the expected sizes. The control reactions with TP8503 all produce the expected products. Sufficient DNA to hybridise with primer 834T (nt 69 to 52 upstream from the first *scrP* codon) is therefore present on the DA26 chromosome,

whereas sufficient DNA to hybridise productively with primer 997M (202–184 nt 28 nt upstream from the first *scrP* codon) is not. The distance between the downstream end of the CAT insertion on the DA26 chromosome and the *Cla* I site in *scrP* is therefore between 0.58 and 0.67 kb. It remains unclear whether the putative *scrP* promoter identified by Smillie [Smillie, 1994] is present on the DA26 chromosome.

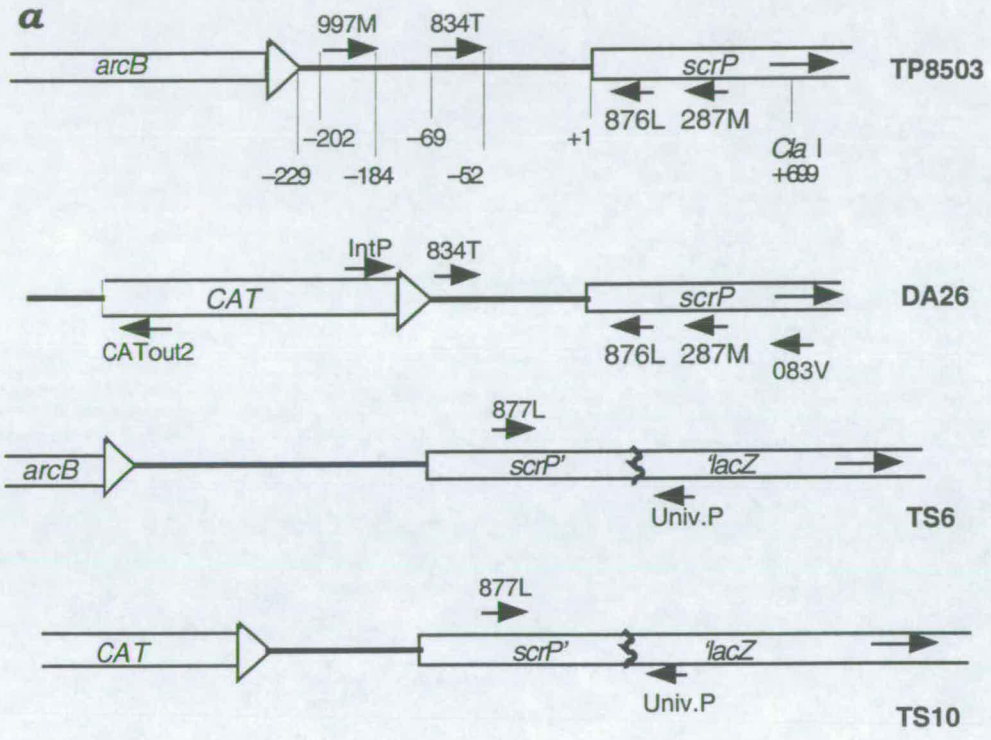


Figure 4.20. Scheme of PCR reactions with TP8503, DA26, TS6 and TS10 cell samples and various primers, as mentioned in the text. Coordinates are relative to translational start for *scrP*. The proposed –35 and –10 elements of a putative *scrP* promoter [Smillie, 1994] lie between –84 and –56 nt upstream from +1.

pSLC3 was the source of the *scrP'–'lacZ* fusion for study in a *ΔarcB* background. However, DA26 is *ΔarcB::*CAT, so the CAT gene in pSLC3 was unsuitable as a marker for fusion insertion. Accordingly, the CAT gene in pSLC3 was first replaced with the Kan^R cassette by digesting the plasmid with *Hind* III and *Bgl* II and ligating the products with the *Hind* III–*Bam*H I fragment (encoding Kan^R) of pUC4K (Fig. 4.22). The ligation products were transformed into

DH5α and the transformants were plated on agar containing X-gal, ampicillin and kanamycin at 30°C. Blue Amp^R Kan^R colonies were purified and the structures of the plasmids present were verified by restriction analysis with *Bam*H I and *Hind* III. the required plasmid was named pSLK3.

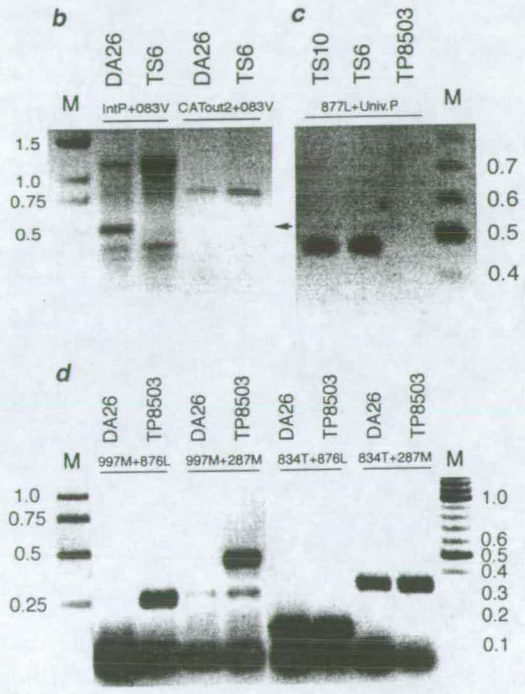


Figure 4.21. PCR products of the reactions (see Fig. 4.20) fractionated on 1% agarose gel. *M*, DNA molecular size markers, molecular sizes given in kb. *a)* PCR of DA26 and TS6 (control) samples with the IntP/083V or CATout2/083V pairs of primers. The specific product of ~0.52 kb present in the DA26 IntP/083V track is marked by an arrow on the right. *b)* – Confirmation of the expected insertion site of *lacZ* on the chromosomes of TS6 and TS10 by PCR with the 877L/Univ.P pair of primers. The expected product of 446 bp was synthesised. TP8503 was used as a control. *c)* – Determination of the region of the CAT gene insertion in DA26 by PCR using primers 997M or 834T as shown in *a*. Expected product sizes for TP8503 are: 997M/876L, 280 bp; 997M/287M, 477 bp; 834T/876L, 147 bp; 834T/287M, 344 bp.

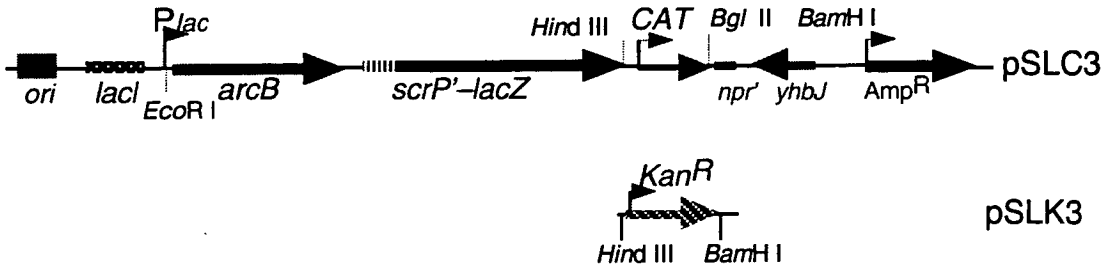


Figure 4.22. Construction of pSLK3 by replacement the CAT cassette in pSLC3 with the Kan^R cassette from pUC4K.

pSLK3 was next used to replace the chromosomal *scrP* with the *scrP'–'lacZ* fusion in MM20*asn*$^+$ (according to the protocol described in section 2.3.5). TP8503 cells were transformed with pSLK3 at 30°C, with Amp^R and Kan^R selection. A purified Amp^R Kan^R transformant was then grown at 30°C in L-broth in the presence of kanamycin only, and a P1 lysate was prepared. This lysate was used to transduce MM20*asn*$^+$ cells at 37°C, selecting for Kan^R. A purified transductant which was Amp^S Kan^R (as well as Cm^S) was named TS9. The structure of its *scrP* region is shown in Fig. 4.12.

The strain TS9 was now transduced with a P1 lysate made on DA26 (*ΔarcB::*CAT) cells and the transductants were plated on L-agar containing X-gal, chloramphenicol, and kanamycin. Despite the short region of homology (about 0.6 kb) available for recombination between the CAT gene and the *scrP'–lacZ* junction (Fig. 4.12), a few Cm^R Kan^R colonies were recovered. All were white, indicating that the level of *scrP–lacZ* expression was very low, in contrast to that in the Kan^R parent TS9, which gave blue colonies on X-gal plates. Note that whereas transcription of *scrP* in TS9 can come from both *arcB* (and possibly other upstream promoters) and the putative *scrP* promoter(s), in the recovered white colonies (they have P_{arcB} deleted) only the putative *scrP* promoters may serve the gene. This suggests that the latter are either inherently very weak, or weakened or inactivated by the *ΔarcB::*CAT replacement. It is also evident that transcription from the CAT promoter is efficiently terminated upstream of *scrP*.

One of the white colonies was purified, and a P1 lysate was prepared on it. TP8503 cells were then transduced with this lysate, selecting for Kan^R and Cm^R. The resulting strain, named TS10, was purified, and its $ArcB^-$ phenotype was confirmed by its dye-sensitivity. The structure of its *scrP* region is shown in Fig. 4.12 and was confirmed by PCR analysis (Fig. 4.20 and 4.21 *b*). The results of the PCR reactions show that amplification with the 877L/Univ.P pair of primers produces a predicted 446 bp product when TS6 and TS10, but not TP8503, samples are used in the reaction.

Due to uncertainty regarding the possible deletion of a putative *scrP* promoter in strain DA26 and its derivatives TS9 and TS10, strain DS941 *arcB*::Tn*5* 5.4 was used as the source of an alternative *arcB* mutation. In this strain all the coding sequence for *arcB* is present, but it is interrupted by the insertion of Tn*5* at codon 473. P1 grown on this strain was used for transduction of TS6 (Fig. 4.12), with selection of Cm^R and Kan^R transductants. The presence of the *arcB*::Tn*5* 5.4 insertion in the resulting purified strain, TS11, was confirmed by dye-sensitivity test.

TS10 and TS11 were assayed for β-galactosidase activity under aerobic as well as under anaerobic conditions. The results are presented in Table 4.5.

In exponential and stationary phases, the amount of β-galactosidase activity in aerobically grown TS10 cells is between 1 and 3 Miller units. It was also detected at a comparable level in anaerobically grown cells. This residual level of ScrP–LacZ synthesis might be due either to slight leakage of transcription from the CAT gene, or due to transcription from the putative *scrP* promoter which may, as discussed above, have wholly or partially escaped deletion from the TS10 chromosome (Fig. 4.12).

TS11 cells, on the other hand, show approximately 3–4 Miller units of β-galactosidase activity in exponential phase and about 5.4

Miller units in stationary phase under aerobic, and 7.7 and 5.7 MU, respectively, under anaerobic conditions.

If we suppose that the *scrP* transcription in TS11 is coming from the putative *scrP* promoter (and not from Tn*5* readthrough), then it can be assumed that about half of *scrP* transcripts produced under aerobic and anaerobic conditions in the normal operon (represented by TS6) come from the independent *scrP* promoter. The lower level of expression in TS10 would then imply that *scrP* promoter is wholly or partially inactivated by the deletion associated with chromosomal *ΔarcB::*CAT replacement in DA26.

This reasoning is based, however, on the results of one or two experiments which need to be repeated. They are also weakened by the fact that in some cases Tn*5*, when inserted into an operon at certain sites, causes low level constitutive expression of distal genes [Berg *et al.*, 1980]. This was assumed to be due to a promoter within ends of Tn*5* which possibly overlaps Tn*5* target sequence boundary.

Discussion

The analysis of expression of genes adjacent to *scrP* showed that they all are transcribed from the promoter upstream of *arcB*, however the study of the expression of *arcB'–'lacZ* and *scrP'–'lacZ* fusions on the chromosome suggested that there might be an independent *scrP* promoter induced under anaerobiosis. Recent work by T. Azam in this laboratory [Azam, 1999] indicated that there is a promoter in the *arcB–scrP* intergenic region, and an even stronger one in this region in the opposite direction whose role is unclear. If so, this promoter, or at least 5' part of it, can be within that part of the intergenic region which is not present in DA26, TS9 and TS10 strain, since these express very little of *scrP'–'lacZ* fusion.

It should be noted that translational β-galactosidase fusions is not the best the choice of study of activity of proteins which are transported through the inner membrane, such as ScrP (a

periplasmic protein, as has been shown in this work) and MtgA. Proteins fused to β–galactosidase might stick in the inner membrane and jam the secretion machinery. β–galactosidase might not fold correctly and consequently does not show activity [Murphy and Beckwith, 1996]. Fortunately for our purposes it was possible to measure the expression of ScrP–LacZ fusion by measuring the activity of the cleaved β-galactosidase.

The MtgA–LacZ fusion protein is more stable therefore it should have difficulties during transportation through the inner membrane, and due to this reason its β-galactosidase activity might not be detected. However, from the overexpression studies it can be concluded that MtgA is expressed many fold lower than ScrP, and therefore MtgA–LacZ fusion expression should be indetectably low for conventional β-galactosidase assays.

The β–galactosidase activity exhibited by ArcB–LacZ fusion protein in this study might have been less than the one produced by the same amount of LacZ protein, due to localisation of *N*-terminal of ArcB–LacZ in certain parts of inner membrane and therefore hindrance of LacZ portion of the fusion to form tetramers necessary for β-galactosidase function.

The *Eco*R I–*Hin*d III DNA fragment present in pFMT1, pTAK1 and other plasmids described further has the potential to encode only the ORFs *arcB*, *scrP* and *mtgA*. The 14 kDa peptide observed by Smillie [Smillie, 1994] in minicell experiment with pFMT1, and in maxicell experiment described in this work, is unaccounted for. It is not seen in the cells overexpressing *scrP* and *mtgA*, therefore it should be assigned to the DNA upstream from *mtgA*. The analysis of the entire *Eco*R I–*Hin*d III region reveals an ORF within *arcB* long enough to encode a polypeptide of 15 kDa. It starts at 972 and finishes at 1374 nt of the *arcB* coding frame. Whether this polypeptide is an artefact of plasmid expression system, or it has a biological significance and is expressed in cells under physiological conditions, remains unclear. Similarity searches for the "15 kDa" ORF retrieve several DNA sequences, all from ongoing bacterial

sequencing projects, and all belonging to γ-subdivision of *Proteobacteria*. Each of these retrieved sequences is a part of a gene corresponding to *E. coli arcB*, and none of them constitute an ORF, as they all contain 1–5 stop codons. Therefore it seems that such a long ORF within *E. coli arcB* is there by chance, and possibly it is not expressed from the chromosome.

References

Aiyar, S.E., Juang, Y.L., Helmann, J.D. & deHaseth, P.L. (1994). Mutations in sigma factor that affect the temperature dependence of transcription from a promoter, but not from a mismatch bubble in double- stranded DNA. *Biochemistry* **33**, 11501–11506.

Alex, L.A., Borkovich, K.A. & Simon, M.I. (1996). Hyphal development in Neurospora crassa: involvement of a two-component histidine kinase. *Proc Natl Acad Sci U S A* **93**, 3416–3421.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.

Appleby, J.L., Parkinson, J.S. & Bourret, R.B. (1996). Signal transduction via the multi-step phosphorelay: not necessarily a road less traveled. *Cell* **86**, 845–848.

Azam, T. (1999). Honours Project. In *ICMB, Division of Biology*. Edinburgh: University of Edinburgh.

Baikalov, I., Schroder, I., Kaczor-Grzeskowiak, M., Grzeskowiak, K., Gunsalus, R.P. & Dickerson, R.E. (1996). Structure of the *Escherichia coli* response regulator NarL. *Biochemistry* **35**, 11053–11061.

Barne, K.A., Bown, J.A., Busby, S.J. & Minchin, S.D. (1997). Region 2.5 of the *Escherichia coli* RNA polymerase sigma70 subunit is responsible for the recognition of the 'extended-10' motif at promoters. *EMBO J* **16**, 4034–4040.

Berg, D.E., Weiss, A. & Crossland, L. (1980). Polarity of Tn*5* insertion mutations in *Escherichia coli*. *J Bacteriol* **142**, 439–446.

Birnboim, H.C. & Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* **7**, 1513–1523.

Blattner, F.R., Plunkett, G.r., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. & Shao, Y. (1997). The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1474.

Bochkarev, A., Pfuetzner, R.A., Edwards, A.M. & Frappier, L. (1997). Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA. *Nature* **385**, 176–181.

Bogachev, A.V., Murtazina, R.A. & Skulachev, V.P. (1993). Cytochrome *d* induction in *Escherichia coli* growing under unfavorable conditions. *FEBS Lett* **336**, 75–78.

Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L. & Boyer, H.W. (1977). Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. *Gene* **2**, 95–113.

Booth, D.R., Sunde, M., Bellotti, V., Robinson, C.V., Hutchinson, W.L., Fraser, P.E., Hawkins, P.N., Dobson, C.M., Radford, S.E., Blake, C.C. & Pepys, M.B. (1997). Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis. *Nature* **385**, 787–793.

Bowers, C.W. & Dombroski, A.J. (1999). A mutation in region 1.1 of sigma70 affects promoter DNA binding by *Escherichia coli* RNA polymerase holoenzyme. *EMBO J* **18**, 709–716.

Bown, J.A., Owens, J.T., Meares, C.F., Fujita, N., Ishihama, A., Busby, S.J. & Minchin, S.D. (1999). Organization of open complexes at *Escherichia coli* promoters. Location of promoter DNA sites close to region 2.5 of the sigma70 subunit of RNA polymerase. *J Biol Chem* **274**, 2263–2270.

Buckle, M., Pemberton, I.K., Jacquet, M.A. & Buc, H. (1999). The kinetics of sigma subunit directed promoter recognition by *E. coli* RNA polymerase. *J Mol Biol* **285**, 955–964.

Burgess, R.R. (1996). Purification of overproduced *Escherichia coli* RNA polymerase sigma factors by solubilizing inclusion bodies and refolding from Sarkosyl. *Methods Enzymol* **273**, 145–149.

Burgess, R.R., Travers, A.A., Dunn, J.J. & Bautz, E.K. (1969). Factor stimulating transcription by RNA polymerase. *Nature* **221**, 43–46.

Buxton, R.S. & Drury, L.S. (1983). Cloning and insertional inactivation of the dye (sfrA) gene, mutation of which affects sex factor F expression and dye sensitivity of *Escherichia coli* K-12. *J Bacteriol* **154**, 1309–1314.

Buxton, R.S. & Drury, L.S. (1984). Identification of the dye gene product, mutational loss of which alters envelope protein composition and also affects sex factor F expression in *Escherichia coli* K-12. *Mol Gen Genet* **194**, 241–247.

Callaci, S., Heyduk, E. & Heyduk, T. (1998). Conformational changes of *Escherichia coli* RNA polymerase sigma70 factor induced by binding to the core enzyme. *J Biol Chem* **273**, 32995–33001.

Callaci, S. & Heyduk, T. (1998). Conformation and DNA binding properties of a single-stranded DNA binding region of sigma 70 subunit from *Escherichia coli* RNA polymerase are modulated by an interaction with the core enzyme. *Biochemistry* **37**, 3312–3320.

Casadaban, M.J. (1976). Transposition and fusion of the lac genes to selected promoters in *Escherichia coli* using bacteriophage lambda and Mu. *J Mol Biol* **104**, 541–555.

Chang, C., Kwok, S.F., Bleecker, A.B. & Meyerowitz, E.M. (1993). Arabidopsis ethylene-response gene ETR1: similarity of product to two-component regulators. *Science* **262**, 539–544.

Chang, H. & Gilbert, W. (1997). A novel zebrafish gene expressed specifically in the photoreceptor cells of the retina. *Biochem Biophys Res Commun* **237**, 84–89.

Chenchik, A.A., Bibilashvili, R., Mirzabekov, A.D. & Shik, V.V. (1982). [Contacts of *Escherichia coli* RNA polymerase subunits with nucleotides of lacUV5 promoter]. *Mol Biol (Mosk)* **16**, 35–46.

Christie, G.E. & Cale, S.B. (1995). The sigma-70 subunit from *Escherichia coli* C differs from that of *E. coli* K-12. *Gene* **162**, 161–162.

Chung, C.T., Niemela, S.L. & Miller, R.H. (1989). One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. *Proc Natl Acad Sci U S A* **86**, 2172–2175.

Close, T.J. & Rodriguez, R.L. (1982). Construction and characterization of the chloramphenicol-resistance gene cartridge: a new approach to the transcriptional mapping of extrachromosomal elements. *Gene* **20**, 305–316.

Colloms, S.D., Alen, C. & Sherratt, D.J. (1998). The ArcA/ArcB two-component regulatory system of *Escherichia coli* is essential for Xer site-specific recombination at psi. *Mol Microbiol* **28**, 521–530.

Compan, I. & Touati, D. (1994). Anaerobic activation of *arcA* transcription in *Escherichia coli*: roles of Fnr and ArcA. *Mol Microbiol* **11**, 955–964.

Craig, M.L., Tsodikov, O.V., McQuade, K.L., Schlax, P.E., Jr., Capp, M.W., Saecker, R.M. & Record, M.T., Jr. (1998). DNA footprints of the two kinetically significant intermediates in formation of an RNA polymerase-promoter open complex: evidence that interactions with start site and downstream DNA induce sequential conformational changes in polymerase and DNA. *J Mol Biol* **283**, 741–756.

Cserzo, M., Wallin, E., Simon, I., von Heijne, G. & Elofsson, A. (1997). Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method. *Protein Eng* **10**, 673–676.

Csonka, L.N. & Clark, A.J. (1980). Construction of an Hfr strain useful for transferring *recA* mutations between *Escherichia coli* strains. *J Bacteriol* **143**, 529–530.

d'Aubenton-Carafa, Y., Brody, E. & Thermes, C. (1990). Prediction of rho-independent *Escherichia coli* transcription terminators. A statistical analysis of their RNA stem-loop structures. *J Mol Biol* **216**, 835–858.

Daniels, D., Zuber, P. & Losick, R. (1990). Two amino acids in an RNA polymerase sigma factor involved in the recognition of adjacent base pairs in the –10 region of a cognate promoter. *Proc Natl Acad Sci U S A* **87**, 8075–8079.

Davis, J.M., Mayor, J. & Plamann, L. (1995). A missense mutation in *rpoD* results in an A-signalling defect in Myxococcus xanthus. *Mol Microbiol* **18**, 943–952.

deHaseth, P.L., Zupancic, M.L. & Record, M.T., Jr. (1998). RNA polymerase-promoter interactions: the comings and goings of RNA polymerase. *J Bacteriol* **180**, 3019–3025.

Denome, S.A., Elf, P.K., Henderson, T.A., Nelson, D.E. & Young, K.D. (1999). *Escherichia coli* mutants lacking all possible combinations of eight penicillin binding proteins: viability, characteristics, and implications for peptidoglycan synthesis. *J Bacteriol* **181**, 3981–3993.

Di Berardino, M., Dijkstra, A., Stuber, D., Keck, W. & Gubler, M. (1996). The monofunctional glycosyltransferase of *Escherichia coli* is a member of a new class of peptidoglycan-synthesising enzymes. *FEBS Lett* **392**, 184–188.

Diederich, B., Tatti, K.M., Jones, C.H., Beall, B. & Moran, C.P., Jr. (1992). Genetic suppression analysis of sigma E interaction with three promoters in sporulating *Bacillus subtilis*. *Gene* **121**, 63–69.

Dombroski, A.J. (1997). Recognition of the –10 promoter sequence by a partial polypeptide of sigma70 *in vitro*. *J Biol Chem* **272**, 3487–3494.

Dombroski, A.J., Johnson, B.D., Lonetto, M. & Gross, C.A. (1996). The sigma subunit of *Escherichia coli* RNA polymerase senses promoter spacing. *Proc Natl Acad Sci U S A* **93**, 8858–8862.

Dombroski, A.J., Walter, W.A. & Gross, C.A. (1993). Amino-terminal amino acids modulate sigma-factor DNA-binding activity. *Genes Dev* **7**, 2446–2455.

Dombroski, A.J., Walter, W.A., Record, M.T., Jr., Siegele, D.A. & Gross, C.A. (1992). Polypeptides containing highly conserved regions of transcription initiation factor sigma 70 exhibit specificity of binding to promoter DNA. *Cell* **70**, 501–512.

Drapal, N. & Sawers, G. (1995). Purification of ArcA and analysis of its specific interaction with the *pfl* promoter-regulatory region. *Mol Microbiol* **16**, 597–607.

Draper, G.C., McLennan, N., Begg, K., Masters, M. & Donachie, W.D. (1998). Only the N-terminal domain of FtsK functions in cell division. *J Bacteriol* **180**, 4621–4627.

Dunn, J.J. & Studier, F.W. (1983). Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *J Mol Biol* **166**, 477–535.

Eichenberger, P., Dethiollaz, S., Buc, H. & Geiselmann, J. (1997). Structural kinetics of transcription activation at the *malT* promoter of *Escherichia coli* by UV laser footprinting. *Proc Natl Acad Sci U S A* **94**, 9022–9027.

Elding, A. (1995). Is *scrP* obligatorily co-transcribed with *arcB* in *Escherichia coli*? In *ICMB*. Edinburgh: University of Edinburgh.

Farewell, A., Diez, A.A., DiRusso, C.C. & Nystrom, T. (1996). Role of the *Escherichia coli* FadR regulator in stasis survival and growth phase-dependent expression of the uspA, fad, and fab genes. *J Bacteriol* **178**, 6443–6450.

Feher, V.A., Zapf, J.W., Hoch, J.A., Whiteley, J.M., McIntosh, L.P., Rance, M., Skelton, N.J., Dahlquist, F.W. & Cavanagh, J. (1997). High-resolution NMR structure and backbone dynamics of the *Bacillus subtilis* response regulator, Spo0F: implications for phosphorylation and molecular recognition. *Biochemistry* **36**, 10015–10025.

Fernandez-Patron, C., Castellanos-Serra, L. & Rodriguez, P. (1992). Reverse staining of sodium dodecyl sulfate polyacrylamide gels by imidazole-zinc salts: sensitive detection of unmodified proteins. *Biotechniques* **12**, 564–573.

Fredrick, K. & Helmann, J.D. (1997). RNA polymerase sigma factor determines start-site selection but is not required for upstream promoter element activation on heteroduplex (bubble) templates. *Proc Natl Acad Sci U S A* **94**, 4982–4987.

Fujita, N., Ishihama, A., Nagasawa, Y. & Ueda, S. (1987). RNA polymerase sigma-related proteins in *Escherichia coli*: detection by antibodies against a synthetic peptide. *Mol Gen Genet* **210**, 5–9.

Georgellis, D., Kwon, O., De Wulf, P. & Lin, E.C. (1998). Signal decay through a reverse phosphorelay in the Arc two-component signal transduction system. *J Biol Chem* **273**, 32864–32869.

Georgellis, D., Lynch, A.S. & Lin, E.C. (1997). *In vitro* phosphorylation study of the arc two-component signal transduction system of *Escherichia coli*. *J Bacteriol* **179**, 5429–5435.

Ghuysen, J.M. (1991). Serine beta-lactamases and penicillin-binding proteins. *Annu Rev Microbiol* **45**, 37–67.

Gopal, V., Ma, H.W., Kumaran, M.K. & Chatterji, D. (1994). A point mutation at the junction of domain 2.3/2.4 of transcription factor sigma 70 abrogates productive transcription and restores its expected mobility on a denaturing gel. *J Mol Biol* **242**, 9–22.

Gross, C.A., Chan, C., Dombroski, A., Gruber, T., Sharp, M., Tupy, J. & Young, B. (1998). The functional and regulatory roles of sigma factors in transcription. *Cold Spring Harb Symp Quant Biol* **63**, 141–155.

Gross, C.A., Lonetto, M. & Losick, R. (1992). Bacterial sigma factors. In *Transcriptional regulation*, pp. 129–176. Edited by S. L. McKnight, Yamamoto, K. R. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

Grossman, A.D., Burgess, R.R., Walter, W. & Gross, C.A. (1983). Mutations in the Ion gene of *E. coli* K12 phenotypically suppress a mutation in the sigma subunit of RNA polymerase. *Cell* **32**, 151–159.

Grossman, A.D., Erickson, J.W. & Gross, C.A. (1984). The *htpR* gene product of *E. coli* is a sigma factor for heat-shock promoters. *Cell* **38**, 383–390.

Guo, Y. & Gralla, J.D. (1998). Promoter opening via a DNA fork junction binding activity. *Proc Natl Acad Sci U S A* **95**, 11655–11660.

Hamada, K., Kato, M., Mizuno, T. & Hakoshima, T. (1999). Crystallographic characterization of a novel protein SixA which exhibits phospho-histidine phosphatase activity in the multistep His– Asp phosphorelay. *Acta Crystallogr D Biol Crystallogr* **55**, 269–271.

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* **166**, 557–580.

Hara, H. & Suzuki, H. (1984). A novel glycan polymerase that synthesizes uncross-linked peptidoglycan in *Escherichia coli*. *FEBS Lett* **168**, 155–160.

Helmann, J.D. & Chamberlin, M.J. (1988). Structure and function of bacterial sigma factors. *Annu Rev Biochem* **57**, 839–872.

Hernandez, V.J. & Cashel, M. (1995). Changes in conserved region 3 of *Escherichia coli* sigma 70 mediate ppGpp-dependent functions *in vivo*. *J Mol Biol* **252**, 536–549.

Hernandez, V.J., Hsu, L.M. & Cashel, M. (1996). Conserved region 3 of *Escherichia coli* final sigma70 is implicated in the process of abortive transcription. *J Biol Chem* **271**, 18775–18779.

Heyduk, E. & Heyduk, T. (1999). Architecture of a complex between the sigma70 subunit of *Escherichia coli* RNA polymerase and the nontemplate strand oligonucleotide. Luminescence resonance energy transfer study. *J Biol Chem* **274**, 3315–3322.

Hillier, L. & al, e. (1995). GenBank accession number H85396. .

Hirokawa, T., Boon-Chieng, S. & Mitaku, S. (1998). SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics* **14**, 378–379.

Hofmann, K. & Stoffel, W. (1993). TMbase — A database of membrane spanning proteins segments. *Biol. Chem. Hoppe-Seyler 166* **347**, 166.

Holtje, J.V. (1998). Growth of the stress-bearing and shape-maintaining murein sacculus of *Escherichia coli*. *Microbiol Mol Biol Rev* **62**, 181–203.

Hu, J.C. & Gross, C.A. (1983). Marker rescue with plasmids bearing deletions in rpoD identifies a dispensable part of *E. coli* sigma factor. *Mol Gen Genet* **191**, 492–498.

Huang, X., Lopez de Saro, F.J. & Helmann, J.D. (1997). Sigma factor mutations affecting the sequence-selective interaction of RNA polymerase with –10 region single-stranded DNA. *Nucleic Acids Res* **25**, 2603–2609.

Igarashi, K. & Ishihama, A. (1991). Bipartite functional map of the *E. coli* RNA polymerase alpha subunit: involvement of the C-terminal region in transcription activation by cAMP–CRP. *Cell* **65**, 1015–1022.

Iost, I. & Dreyfus, M. (1995). The stability of *Escherichia coli lacZ* mRNA depends upon the simultaneity of its synthesis and translation. *EMBO J* **14**, 3252–3261.

Ishige, K., Nagasawa, S., Tokishita, S. & Mizuno, T. (1994). A novel device of bacterial signal transducers. *EMBO J* **13**, 5195–5202.

Iuchi, S. (1993). Phosphorylation/dephosphorylation of the receiver module at the conserved aspartate residue controls transphosphorylation activity of histidine kinase in sensor protein ArcB of *Escherichia coli. J Biol Chem* **268**, 23972–23980.

Iuchi, S., Cameron, D.C. & Lin, E.C. (1989a). A second global regulator gene *(arcB)* mediating repression of enzymes in aerobic pathways of *Escherichia coli. J Bacteriol* **171**, 868–873.

Iuchi, S., Chepuri, V., Fu, H.A., Gennis, R.B. & Lin, E.C. (1990b). Requirement for terminal cytochromes in generation of the aerobic signal for the arc regulatory system in *Escherichia coli*: study utilizing deletions and lac fusions of cyo and cyd. *J Bacteriol* **172**, 6020–6025.

Iuchi, S., Furlong, D. & Lin, E.C. (1989b). Differentiation of *arcA*, *arcB*, and *cpxA* mutant phenotypes of *Escherichia coli* by sex pilus formation and enzyme regulation. *J Bacteriol* **171**, 2889–2893.

Iuchi, S. & Lin, E.C. (1988). *arcA (dye)*, a global regulatory gene in *Escherichia coli* mediating repression of enzymes in aerobic pathways. *Proc Natl Acad Sci U S A* **85**, 1888–1892.

Iuchi, S. & Lin, E.C. (1992b). Mutational analysis of signal transduction by ArcB, a membrane sensor protein responsible for anaerobic repression of operons involved in the central aerobic pathways in *Escherichia coli. J Bacteriol* **174**, 3972–3980.

Iuchi, S. & Lin, E.C. (1992a). Purification and phosphorylation of the Arc regulatory components of *Escherichia coli. J Bacteriol* **174**, 5617–5623.

Iuchi, S. & Lin, E.C. (1993). Adaptation of *Escherichia coli* to redox environments by gene expression. *Mol Microbiol* **9**, 9–15.

Iuchi, S., Matsuda, Z., Fujiwara, T. & Lin, E.C. (1990a). The arcB gene of *Escherichia coli* encodes a sensor-regulator protein for anaerobic repression of the *arc* modulon. *Mol Microbiol* **4**, 715–727.
modulon. *Mol Microbiol* **4**, 715–727.

Jenkins, A.J., March, J.B., Oliver, I.R. & Masters, M. (1986). A DNA fragment containing the groE genes can suppress mutations in the *Escherichia coli dnaA* gene. *Mol Gen Genet* **202**, 446–454.

Jones, C.H. & Moran, C.P., Jr. (1992). Mutant sigma factor blocks transition between promoter binding and initiation of transcription. *Proc Natl Acad Sci U S A* **89**, 1958–1962.

Jones, C.H., Tatti, K.M. & Moran, C.P., Jr. (1992). Effects of amino acid substitutions in the −10 binding region of sigma E from *Bacillus subtilis. J Bacteriol* **174**, 6815–6821.

Joo, D.M., Ng, N. & Calendar, R. (1997). A sigma32 mutant with a single amino acid change in the highly conserved region 2.2 exhibits reduced core RNA polymerase affinity. *Proc Natl Acad Sci U S A* **94**, 4907–4912.

Joo, D.M., Nolte, A., Calendar, R., Zhou, Y.N. & Jin, D.J. (1998). Multiple regions on the *Escherichia coli* heat shock transcription factor sigma32 determine core RNA polymerase binding specificity. *J Bacteriol* **180**, 1095–1102.

Juang, Y.L. & Helmann, J.D. (1994). A promoter melting region in the primary sigma factor of *Bacillus subtilis*. Identification of functionally important aromatic amino acids. *J Mol Biol* **235**, 1470–1488.

Juang, Y.L. & Helmann, J.D. (1995). Pathway of promoter melting by *Bacillus subtilis* RNA polymerase at a stable RNA promoter: effects of temperature, delta protein, and sigma factor mutations [published erratum appears in Biochemistry 1995 Oct 31;34(43):14270]. *Biochemistry* **34**, 8465–8473.
Kakimoto, T. (1996). CKI1, a histidine kinase homolog implicated in cytokinin signal transduction. *Science* **274**, 982–985.

Kato, M., Mizuno, T., Shimizu, T. & Hakoshima, T. (1997). Insights into multistep phosphorelay from the crystal structure of the C-terminal HPt domain of ArcB. *Cell* **88**, 717–723.

Kenney, T.J., York, K., Youngman, P. & Moran, C.P., Jr. (1989). Genetic evidence that RNA polymerase associated with sigma A factor uses a sporulation-specific promoter in *Bacillus subtilis*. *Proc Natl Acad Sci U S A* **86**, 9109–9113.

Khattar, M.M., Begg, K.J. & Donachie, W.D. (1994). Identification of FtsW and characterization of a new ftsW division mutant of *Escherichia coli*. *J Bacteriol* **176**, 7140–7147.

Klenk, H.P., Clayton, R.A., Tomb, J.F., White, O., Nelson, K.E., Ketchum, K.A., Dodson, R.J., Gwinn, M., Hickey, E.K., Peterson, J.D., Richardson, D.L., Kerlavage, A.R., Graham, D.E., Kyrpides, N.C., Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S., Kirkness, E.F., Dougherty, B.A., McKenney, K., Adams, M.D., Loftus, B., Venter, J.C. & et al. (1997). The complete genome sequence of the hyperthermophilic, sulphate- reducing archaeon *Archaeoglobus fulgidus* [published erratum appears in Nature 1998 Jul 2;394(6688):101]. *Nature* **390**, 364–370.

Kohara, Y., Akiyama, K. & Isono, K. (1987). The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. *Cell* **50**, 495–508.

Kubori, T. & Shimamoto, N. (1996). A branched pathway in the early stage of transcription by *Escherichia coli* RNA polymerase. *J Mol Biol* **256**, 449–457.

Kumar, A., Grimes, B., Logan, M., Wedgwood, S., Williamson, H. & Hayward, R.S. (1995b). A hybrid sigma subunit directs RNA polymerase to a hybrid promoter in *Escherichia coli*. *J Mol Biol* **246**, 563–571.

Kumar, A., Malloch, R.A., Fujita, N., Smillie, D.A., Ishihama, A. & Hayward, R.S. (1993). The minus 35-recognition region of *Escherichia coli* sigma 70 is inessential for initiation of transcription at an "extended minus 10" promoter. *J Mol Biol* **232**, 406–418.

Kumar, A., Williamson, H.S., Fujita, N., Ishihama, A. & Hayward, R.S. (1995a). A partially functional 245-amino-acid internal deletion derivative of *Escherichia coli* sigma 70. *J Bacteriol* **177**, 5193–5196.

Kunkel, T.A. (1985). Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc Natl Acad Sci U S A* **82**, 488–492.

Kyte, J. & Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157**, 105–132.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685.

Lafreniere, R.G., Rochefort, D.L., Kibar, Z., Fon, E.A., Han, F., Cochius, J., Kang, X., Baird, S., Korneluk, R.G., Andermann, E., Rommens, J.M., Rouleau, G.A. (1996). Isolation and characterization of GT335, a novel human gene conserved in *Escherichia coli* and mapping to 21q22.3 *Genomics* **38**, 264–272.

Lesley, S.A. & Burgess, R.R. (1989). Characterization of the *Escherichia coli* transcription factor sigma 70: localization of a region involved in the interaction with core RNA polymerase. *Biochemistry* **28**, 7728–7734.

Liebke, H., Gross, C., Walter, W. & Burgess, R. (1980). A new mutation *rpoD800*, affecting the sigma subunit of *E. coli* RNA polymerase is allelic to two other sigma mutants. *Mol Gen Genet* **177**, 277–282.

Lonetto, M., Gribskov, M. & Gross, C.A. (1992). The sigma 70 family: sequence conservation and evolutionary relationships. *J Bacteriol* **174**, 3843–3849.

Lowe, P.A., Aebi, U., Gross, C. & Burgess, R.R. (1981). *In vitro* thermal inactivation of a temperature-sensitive sigma subunit mutant (rpoD800) of *Escherichia coli* RNA polymerase proceeds by aggregation. *J Biol Chem* **256**, 2010–2015.

Lynch, A.S., Lin, E. C. C. (1996a). Responses to molecular oxygen. In *Escherichia coli and Salmonella typhimirium*, pp. 1526–1538. Edited by Neidhardt *et al.* Washington D. C.: ASM Press.

Lynch, A.S. & Lin, E.C. (1996b). Transcriptional control mediated by the ArcA two-component response regulator protein of *Escherichia coli*: characterization of DNA binding at target promoters. *J Bacteriol* **178**, 6238–6249.

Malhotra, A., Severinova, E. & Darst, S.A. (1996). Crystal structure of a sigma 70 subunit fragment from *E. coli* RNA polymerase. *Cell* **87**, 127–136.

Marr, M.T. & Roberts, J.W. (1997). Promoter recognition as measured by binding of polymerase to nontemplate strand oligonucleotide. *Science* **276**, 1258–1260.

Matsushika, A. & Mizuno, T. (1998b). A dual-signaling mechanism mediated by the ArcB hybrid sensor kinase containing the histidine-containing phosphotransfer domain in *Escherichia coli*. *J Bacteriol* **180**, 3973–3977.

Matsushika, A. & Mizuno, T. (1998a). The structure and function of the histidine-containing phosphotransfer (HPt) signaling domain of the *Escherichia coli* ArcB sensor. *J Biochem (Tokyo)* **124**, 440–445.

McLaren, R.S., Newbury, S.F., Dance, G.S., Causton, H.C. & Higgins, C.F. (1991). mRNA degradation by processive 3'–5' exoribonucleases *in vitro* and the implications for prokaryotic mRNA decay *in vivo*. *J Mol Biol* **221**, 81–95.

Merrick, M., Jones, D.H. & Thomas, C.M. (1993). Location of the *rpoN* gene on the physical map of *Escherichia coli*. *J Bacteriol* **175**, 1548–1549.

Miller, J. (1982). *Experiments in molecular genetics*. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

Mizuno, T. (1997). Compilation of all genes encoding two-component phosphotransfer signal transducers in the genome of *Escherichia coli*. *DNA Res* **4**, 161–168.

Murphy, C.K. & Beckwith, J. (1996). Export of proteins to the cell envelope in *Escherichia coli*. In *Escherichia coli and Salmonella typhimirium*, pp. p.967–977. Edited by Neidhardt *et al.* Washington D. C.: ASM Press.

Nagai, H. & Shimamoto, N. (1997). Regions of the *Escherichia coli* primary sigma factor sigma70 that are involved in interaction with RNA polymerase core enzyme. *Genes Cells* **2**, 725–734.

Nagai, K. (1997–1999). PSORT: http://psort.nibb.ac.jp:8800/.

Nagamine, K., Kudoh, J., Minoshima, S., Kawasaki, K., Asakawa, S., Ito, F. & Shimizu, N. (1996). Isolation of cDNA for a novel human protein KNP-I that is homologous to the *E. coli* SCRP-27A protein from the autoimmune polyglandular disease type I (APECED) region of chromosome 21q22.3. *Biochem Biophys Res Commun* **225**, 608–616.

Nakagawa, J., Tamaki, S., Tomioka, S. & Matsuhashi, M. (1984). Functional biosynthesis of cell wall peptidoglycan by polymorphic bifunctional polypeptides. Penicillin-binding protein 1Bs of *Escherichia coli* with activities of transglycosylase and transpeptidase. *J Biol Chem* **259**, 13937–13946.

Narayanan, C.S. & Krakow, J.S. (1983). Chemical modifications of the sigma subunit of the *E. coli* RNA polymerase. *Nucleic Acids Res* **11**, 2701–2716.

Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng* **10**, 1–6.

Ogino, T., Matsubara, M., Kato, N., Nakamura, Y. & Mizuno, T. (1998). An *Escherichia coli* protein that exhibits phosphohistidine phosphatase activity towards the HPt domain of the ArcB sensor involved in the multistep His-Asp phosphorelay. *Mol Microbiol* **27**, 573–585.

Ota, I.M. & Varshavsky, A. (1993). A yeast protein similar to bacterial two-component regulators. *Science* **262**, 566–569.

Owens, J.T., Chmura, A.J., Murakami, K., Fujita, N., Ishihama, A. & Meares, C.F. (1998a). Mapping the promoter DNA sites proximal to conserved regions of sigma 70 in an *Escherichia coli* RNA polymerase-*lac*UV5 open promoter complex. *Biochemistry* **37**, 7670–7675.

Owens, J.T., Miyake, R., Murakami, K., Chmura, A.J., Fujita, N., Ishihama, A. & Meares, C.F. (1998b). Mapping the sigma70 subunit contact sites on *Escherichia coli* RNA polymerase with a sigma70-conjugated chemical protease. *Proc Natl Acad Sci U S A* **95**, 6021–6026.

Park, W. & Matsuhashi, M. (1984). Staphylococcus aureus and Micrococcus luteus peptidoglycan transglycosylases that are not penicillin-binding proteins. *J Bacteriol* **157**, 538–544.

Park, W., Seto, H., Hakenbeck, R. & Matsuhashi, M. (1985). Major peptidoglycan transglycosylase activity in *Streptococcus Pneumoniae* that is not a penicillin-binding protein. *FEMS Microbiology Letters* **27**, 45–48.

Parkinson, J.S. & Kofoid, E.C. (1992). Communication modules in bacterial signaling proteins. *Annu Rev Genet* **26**, 71–112.

Pogliano, J., Lynch, A.S., Belin, D., Lin, E.C. & Beckwith, J. (1997). Regulation of *Escherichia coli* cell envelope proteins involved in protein folding and degradation by the Cpx two-component system. *Genes Dev* **11**, 1169–1182.

Polyakov, A., Severinova, E. & Darst, S.A. (1995). Three-dimensional structure of *E. coli* core RNA polymerase: promoter binding and elongation conformations of the enzyme. *Cell* **83**, 365–373.

Powell, B.S., Court, D.L., Inada, T., Nakamura, Y., Michotey, V., Cui, X., Reizer, A., Saier, M.H., Jr. & Reizer, J. (1995). Novel proteins of the phosphotransferase system encoded within the rpoN operon of *Escherichia coli*. Enzyme IIANtr affects growth on organic nitrogen and the conditional lethality of an erats mutant. *J Biol Chem* **270**, 4822–4839.

Roberts, C.W. & Roberts, J.W. (1996). Base-specific recognition of the nontemplate strand of promoter DNA by *E. coli* RNA polymerase. *Cell* **86**, 495–501.

Rong, J.C. & Helmann, J.D. (1994). Genetic and physiological studies of *Bacillus subtilis* sigma A mutants defective in promoter melting. *J Bacteriol* **176**, 5218–5224.

Rost, B., Casadio, R., Fariselli, P. & Sander, C. (1995). Transmembrane helices predicted at 95% accuracy. *Protein Sci* **4**, 521–533.

Rudolph, J. & Oesterhelt, D. (1995). Chemotaxis and phototaxis require a CheA histidine kinase in the archaeon Halobacterium salinarium. *EMBO J* **14**, 667–673.

Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989). *Molecular cloning: a laboratory manual*, 2nd edn. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

Sancar, A. & Rupert, C.S. (1978). Determination of plasmid molecular weights from ultraviolet sensitivities. *Nature* **272**, 471–472.

Schagger, H. & von Jagow, G. (1987). Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal Biochem* **166**, 368–379.

Schaubach, O.L. & Dombroski, A.J. (1999). Transcription initiation at the flagellin promoter by RNA polymerase carrying sigma28 from *Salmonella typhimurium*. *J Biol Chem* **274**, 8757–8763.

Schiffer, G., Templin, M.F. & Holtje, J.-V. (1997). Cloning and biochemical characterisation of the bifunctional penicillin-binding protein 1C from *Escherichia coli*. GenBank accession No. U88571. .

Schuster, S.S., Noegel, A.A., Oehme, F., Gerisch, G. & Simon, M.I. (1996). The hybrid histidine kinase DokA is part of the osmotic response system of Dictyostelium. *EMBO J* **15**, 3880–3889.

Scott, H.S., Chen, H., Rossier, C., Lalioti, M.D., Antonarakis, S.E., (1997). Isolation of a human gene (HES1) with homology to an *Escherichia coli* and a zebrafish protein that maps to chromosome 21q22.3. *Hum Genet* **99**, 616–23

Sen, R., Nagai, H., Hernandez, V.J. & Shimamoto, N. (1998). Reduction in abortive transcription from the lambdaPR promoter by mutations in region 3 of the sigma70 subunit of *Escherichia coli* RNA polymerase. *J Biol Chem* **273**, 9872–9877.

Severinov, K., Fenyo, D., Severinova, E., Mustaev, A., Chait, B.T., Goldfarb, A. & Darst, S.A. (1994). The sigma subunit conserved region 3 is part of "5'-face" of active center of *Escherichia coli* RNA polymerase. *J Biol Chem* **269**, 20826–20828.

Severinova, E., Severinov, K., Fenyo, D., Marr, M., Brody, E.N., Roberts, J.W., Chait, B.T. & Darst, S.A. (1996). Domain organization of the *Escherichia coli* RNA polymerase sigma 70 subunit. *J Mol Biol* **263**, 637–647.

Shimamoto, N., Kasciukovich, T., Nagai, H. & Hayward, R. (1998). Efficient solubilisation of proteins overproduced as inclusion bodies by use of an extreme concentration of glycerol. *Technical Tips Online*, T01576.

Shorenstein, R.G. & Losick, R. (1973). Comparative size and properties of the sigma subunits of ribonucleic acid polymerase from *Bacillus subtilis* and *Escherichia coli*. *J Biol Chem* **248**, 6170–6173.

Siebenlist, U., Simpson, R.B. & Gilbert, W. (1980). *E. coli* RNA polymerase interacts homologously with two different promoters. *Cell* **20**, 269–281.

Siegele, D.A., Hu, J.C., Walter, W.A. & Gross, C.A. (1989). Altered promoter recognition by mutant forms of the sigma 70 subunit of *Escherichia coli* RNA polymerase. *J Mol Biol* **206**, 591–603.

Silverman, P.M., Rother, S. & Gaudin, H. (1991a). Arc and Sfr functions of the *Escherichia coli* K-12 arcA gene product are genetically and physiologically separable. *J Bacteriol* **173**, 5648–5652.

Silverman, P.M., Wickersham, E. & Harris, R. (1991b). Regulation of the F plasmid *traY* promoter in *Escherichia coli* by host and plasmid factors. *J Mol Biol* **218**, 119–128.

Simpson, R.B. (1979). The molecular topography of RNA polymerase-promoter interaction. *Cell* **18**, 277–285.

Smillie, D.A. (1994). Genes encoding sigma cross-reacting proteins in *Escherichia coli.* In *Institute of Cell and Molecular Biology.* Edinburgh: University of Edinburgh.

Sonnhammer, E.L., von Heijne, G. & Krogh, A. (1998). A hidden Markov model for predicting transmembrane helices in protein sequences. *Ismb* **6**, 175–182.

Spratt, B.G., Zhou, J., Taylor, M. & Merrick, M.J. (1996). Monofunctional biosynthetic peptidoglycan transglycosylases. *Mol Microbiol* **19**, 639–640.

Stock, J.B., Surette, M.G., Levit, M. & Park, P. (1995). Two-component signal transduction systems: structure–function relationships and mechanisms of catalysis. In *Two-Component Signal Transduction*, pp. 25—51. Edited by Hoch and Silhavy. Washington, D. C.: ASM Press.

Stoker, N.G., Pratt, J.M. & Spratt, B.G. (1983). Identification of the *rodA* gene product of *Escherichia coli. J Bacteriol* **155**, 854–859.

Strohmaier, H., Noiges, R., Kotschan, S., Sawers, G., Hogenauer, G., Zechner, E.L. & Koraimann, G. (1998). Signal transduction and bacterial conjugation: characterization of the role of ArcA in regulating conjugative transfer of the resistance plasmid R1. *J Mol Biol* **277**, 309–316.

Studier, F.W. & Moffatt, B.A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* **189**, 113–130.

Studier, F.W., Rosenberg, A.H., Dunn, J.J. & Dubendorff, J.W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol* **185**, 60–89.

Tabor, S. & Richardson, C.C. (1985). A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proc Natl Acad Sci U S A* **82**, 1074–1078.

Taku, A., Stuckey, M. & Fan, D.P. (1982). Purification of the peptidoglycan transglycosylase of Bacillus megaterium. *J Biol Chem* **257**, 5018–5022.

Tatti, K.M., Jones, C.H. & Moran, C.P., Jr. (1991). Genetic evidence for interaction of sigma E with the *spoIIID* promoter in *Bacillus subtilis. J Bacteriol* **173**, 7828–7833.

Tatti, K.M. & Moran, C.P., Jr. (1995). sigma E changed to sigma B specificity by amino acid substitutions in its –10 binding region. *J Bacteriol* **177**, 6506–6509.

Taylor, B.L. & Zhulin, I.B. (1999). PAS domains: internal sensors of oxygen, redox potential, and light. *Microbiol Mol Biol Rev* **63**, 479–506.

Taylor, K.L., Cheng, N., Williams, R.W., Steven, A.C. & Wickner, R.B. (1999). Prion domain initiation of amyloid formation *in vitro* from native Ure2p. *Science* **283**, 1339–1343.

Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.

Townsley, F.M. (1991). Honours Project. In *ICMB*. Edinburgh: University of Edinburgh.

Traviglia, S.L., Datwyler, S.A. & Meares, C.F. (1999). Mapping protein-protein interactions with a library of tethered cutting reagents: the binding site of sigma 70 on *Escherichia coli* RNA polymerase. *Biochemistry* **38**, 4259–4265.

Tsuzuki, M., Ishige, K. & Mizuno, T. (1995). Phosphotransfer circuitry of the putative multi-signal transducer, ArcB, of *Escherichia coli*: *in vitro* studies with mutants. *Mol Microbiol* **18**, 953–962.

Tusnady, G.E. & Simon, I. (1998). Principles governing amino acid composition of integral membrane proteins: application to topology prediction. *J Mol Biol* **283**, 489–506.

Ueshima, R., Fujita, N. & Ishihama, A. (1992). Identification of *Escherichia coli* proteins cross-reacting with antibodies against region 2.2 peptide of RNA polymerase sigma subunit. *Biochem Biophys Res Commun* **184**, 634–639.

Vieira, J. & Messing, J. (1982). The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* **19**, 259–268.

Volkman, B.F., Nohaile, M.J., Amy, N.K., Kustu, S. & Wemmer, D.E. (1995). Three-dimensional solution structure of the N-terminal receiver domain of NTRC. *Biochemistry* **34**, 1413–1424.

Vollmer, W., von Rechenberg, M. & Holtje, J.V. (1999). Demonstration of molecular interactions between the murein polymerase PBP1B, the lytic transglycosylase MltA, and the scaffolding protein MipA of *Escherichia coli*. *J Biol Chem* **274**, 6726–6734.

Volz, K. (1995). Structural and functional conservation in response regulators. In *Two-Component Signal Transduction*, pp. 53—64. Edited by Hoch and Silhavy. Washington, D. C.: ASM Press.

Waldburger, C., Gardella, T., Wong, R. & Susskind, M.M. (1990). Changes in conserved region 2 of *Escherichia coli* sigma 70 affecting promoter recognition. *J Mol Biol* **215**, 267–276.

Waldburger, C. & Susskind, M.M. (1994). Probing the informational content of *Escherichia coli* sigma 70 region 2.3 by combinatorial cassette mutagenesis. *J Mol Biol* **235**, 1489-1500.

Wang, N., Shaulsky, G., Escalante, R. & Loomis, W.F. (1996). A two-component histidine kinase gene that functions in Dictyostelium development. *EMBO J* **15**, 3890–3898.

Wellman, A. & Meares, C.F. (1991). Footprint of the sigma protein: a re-examination. *Biochem Biophys Res Commun* **177**, 140–144.

Whittaker, P.A., Campbell, A.J., Southern, E.M. & Murray, N.E. (1988). Enhanced recovery and restriction mapping of DNA fragments cloned in a new lambda vector. *Nucleic Acids Res* **16**, 6725–6736.

Wilkinson, J.Q., Lanahan, M.B., Yen, H.C., Giovannoni, J.J. & Klee, H.J. (1995). An ethylene-inducible component of signal transduction encoded by never-ripe [see comments]. *Science* **270**, 1807–1809.

Wilson, C. & Dombroski, A.J. (1997). Region 1 of sigma70 is required for efficient isomerization and initiation of transcription by *Escherichia coli* RNA polymerase. *J Mol Biol* **267**, 60–74.

Wright, J.J. & Hayward, R.S. (1987). Transcriptional termination at a fully rho-independent site in *Escherichia coli* is prevented by uninterrupted translation of the nascent RNA. *EMBO J* **6**, 1115–1119.

Yaku, H., Kato, M., Hakoshima, T., Tsuzuki, M. & Mizuno, T. (1997). Interaction between the CheY response regulator and the histidine- containing phosphotransfer (HPt) domain of the ArcB sensory kinase in *Escherichia coli. FEBS Lett* **408**, 337–340.

Yousif, S.Y., Broome-Smith, J.K. & Spratt, B.G. (1985). Lysis of *Escherichia coli* by beta-lactam antibiotics: deletion analysis of the role of penicillin-binding proteins 1A and 1B. *J Gen Microbiol* **131**, 2839.

Yura T, Ishihama A. (1979). Genetics of bacterial RNA polymerases. *Annu Rev Genet* **13**, 59–97.

Zhou, Y.N., Walter, W.A. & Gross, C.A. (1992). A mutant sigma 32 with a small deletion in conserved region 3 of sigma has reduced affinity for core RNA polymerase. *J Bacteriol* **174**, 5005–5012.

Zuber, P., Healy, J., Carter, H.L.d., Cutting, S., Moran, C.P., Jr. & Losick, R. (1989). Mutation changing the specificity of an RNA polymerase sigma factor. *J Mol Biol* **206**, 605–614.

Technical Tips Online

Simple History Results

Efficient solubilization of proteins overproduced as inclusion bodies by use of an extreme concentration of glycerol

Nobuo Shimamoto [a] nshimamoto@labstrg-1.lab.nig.ac.jp, Taciana Kasciukovich [c], Hiroki Nagai [b] and Richard S. Hayward [d]

[a]Nobuo Shimamoto is at the National Institute of Genetics, Structural Biology Center, Yata, Mishima, Japan 411-8540.[b]Hiroki Nagai is at the National Institute of Genetics, Structural Biology Center, Yata, Mishima, Japan 411-8540.[c]Taciana Kasciukovich is at the Institute of Cell and Molecular Biology, The University of Edinburgh, Mayfield Road, Edinburgh, UK EH9 3JR.[d]Richard S. Hayward is at the Institute of Cell and Molecular Biology, The University of Edinburgh, Mayfield Road, Edinburgh, UK EH9 3JR.

Related fulltext articles on BioMedNet

Search
Latest Issue
Browse Issues
Browse Categories
About TTO

Quick Site Search
GO
BioMedNet
Advanced site search

Questionnaire
Give us your opinion...

Article Outline

Protocol
Solubilization of inclusion bodies in guanidine hydrochloride (GnHCl) solution and removal of GnHCl
Dilution in an aqueous buffer
Results and discussion
Acknowledgements
References
Glossary
Copyright

Create new comment

Proteins overproduced in *Escherichia coli* sometimes form inclusion bodies, which can then be dissolved with the help of chaotropic reagents such as guanidine hydrochloride or urea. To recover proteins in aqueous buffers suitable for biochemical use, the concentrations of chaotropic reagents then must be reduced, usually by dialysis or rapid dilution (Ref. 1, 2, 3). Unwanted re-precipitation often happens at this stage. Here, we report that dialysis against buffers containing 75% glycerol prevents such re-precipitation, and that the solubility can be efficiently maintained through a subsequent rapid dilution of the glycerol.

Protocol

The solubilization procedure consisted of two steps:

Solubilization of inclusion bodies in guanidine hydrochloride (GnHCl) solution and removal of GnHCl

The inclusion bodies are conventionally prepared from *E. coli* cell homogenates by centrifugation for 30 min at 13 000 rpm at 4°C, and suspended in TEM buffer (10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 1 mM 2-mercaptoethanol) containing GnHCl. The final concentration of GnHCl should not be less than 6M. Any insoluble materials are removed by centrifugation under the same conditions. The supernatant is then rapidly diluted with 10 volumes of a buffer containing no denaturants, and kept for more than 20 min at 4°C, after which the suspension is centrifuged again. The resulting precipitate should contain proteins that are difficult to solubilize by ordinary procedures. The solubilization, dilution and centrifugation cycle is repeated once more, completing a very efficient purification step for such proteins.

The washed precipitates are again dissolved in TEM buffer containing 6M GnHCl in a graduated conical tube, at a concentration of 2–20 mg protein/ml, as judged by the BCA protein assay (Pierce, Rockford). The solution is transferred into dialysis tubing and an equal volume of glycerol is added to prevent gelation of proteins during dialysis. After sealing the tubing, the solution is mixed by inversion. The total protein concentration at this stage should be less than 10 mg/ml; preferably less than 2 mg/ml. If the tubing contains less than 5 ml, it is placed in a 50 ml conical tube containing 45 ml of a 3:1 (v/v) mixture of glycerol and TEM buffer. If the volume is greater, a 300–1000 ml centrifuge bottle is used. The tube or bottle is then rotated at 10 rpm and 4°C for more than 10 h. The 75% glycerol buffer is changed twice each 2–4 h or once after 8 h. The dialyzed solution can then be stored in a freezer at −20°C.

Dilution in an aqueous buffer

The 75% glycerol solution containing 1–10 mg protein/ml is rapidly diluted with 6.5–14 volumes of TEM buffer containing 0.1 M NaCl to give a final solution containing 10–5% glycerol. For large scale preparations, the glycerol solution is added dropwise at the center of the diluent, which is simultaneously subjected to vigorous magnetic stirring. For small scale preparations, drops of the glycerol solution are placed on the wall or lid of an Eppendorf tube containing diluent, followed by Vortex mixing at full-speed for 10 sec. The diluted samples are placed on ice for more than 10 min, then centrifuged at not less than 10 000?g to remove possible precipitate.

Results and discussion

No precipitation was observed with any of the ten test proteins in the first step (Table 1). The high glycerol concentration is essential. Significant precipitation was observed when it was decreased to 50%. In the second step, rapid dilution is required for a high yield. The second step could be further improved by selecting better buffer components and by use of ion-exchange resin during renaturation (Ref. 1, 2, 3), but we here used only TEM buffer for dilution.

Table 1. Summary of solubilization of proteins in the form of inclusion bodies.

#	Source	Protein/modification[a]	Vector/host	Solubilized[b]	Activity[c]	Improvement[d]
1	*Escherichia coli*	Sigma-70 with deletion of aa 130 through 374 and carrying an HMKtag[e]. Also 4 derivatives[f]	pTrc99A[g]/C600	>95%	+	10
				>95%[g]	Not determined	10
2	*Drosophila melanogaster*	175 aa section of TFIIH homologue fused with GST at N-terminus	pGEX-4T-2[h]/BL21(DE3)[i]	9%	+	180
3	*Arabidopsis thaliana*	Chloroplast SigA with signal peptide	pET15b[i]/BL21(DE3, pLysS)[I]	60%	Not determined	85
		ibid . without signal peptide	ibid	15%	Not determined	11
4	*Tetrahymena thermophila*	A translational factor $(His)_6$ at C-terminus	pTE3[i]/BL21(DE3)[i]	>95%	Not determined	110
5	Hen egg	Avidin[j] with a saturating modification by FITC[k]	-	>95%	+	10

[a]Prepared in this study (1, 5) or gifts from Dr Hiroyuki Kose in the National Institute of Genetics (2), or from Dr Kan Tanaka (3) and Dr Koichi Ito (4) of Tokyo University.

[b]Fractions of proteins recovered in the supernatant of the final buffer containing 5% glycerol.

[c]Activities were measured by transcription initiation (1), binding to a glutathione column (2), and binding to biotin (5).

[d]Ratios of the yields by the new procedure and the control procedure.

[e]See Ref. 8

[f]Derivatives with single amino-acid substitutions.

[g]The K502C mutant of the sigma-70 derivative (50 μg) was refolded on 0.1 ml of DEAE Toyopearl S650 resin (Toyobo).

[h]From Amersham Pharmacia Biotech

[i]From Novagen

[j]From Taiyo Kagaku, Tokyo

[k]After modification with FITC (Molecular Probe), unreacted FITC was removed by adding acetone to 50%, with precipitation of the modified avidin.

In the conventional method that we have used as a standard for comparison, 0.1–1.0 mg total protein/ml in 6M GnHCl solution was dialyzed directly against TEM buffer containing 0.1 M NaCl and 5% glycerol, and the dialyzed solution was centrifuged to remove insoluble proteins. The initial protein concentrations used in the new procedure were 15-fold more concentrated, to compensate for the 15-fold rapid dilution in the second step. The soluble fractions obtained by the two procedures were analyzed by SDS PAGE and the gels were reverse-stained (Ref. 4, 5). Proteins were quantitated by reference to bands of various amounts of 6M GnHCl solutions or 75% glycerol solutions. The new procedure improved yields by 10 to 180 fold (Table 1). Two further procedures were tested, using one of # 1 proteins in Table 1. A stepwise dialysis procedure (3 M, 1.5 M, 0.75 M, 0.4 M, and 0 M GnHCl for 3 h each) gave the same yield as the conventional dialysis. By contrast, a resin-assisted version of the new procedure, which is more convenient for purification, gave an equally excellent result (Table 1). An ion-exchange resin was added in the tubing, and dialyzed against the 75% and then 5% glycerol buffers in place of rapid dilution. The resin was collected by centrifugation, and the protein was eluted with TEM buffer containing 0.5M KCl.

The new procedure was applicable to inclusion bodies formed by proteins with intact primary structures (# 3), deletion derivatives (# 1) or artificial structures (# 2, 4), as well as a chemically modified protein (# 5). Activities of some proteins were confirmed, indicating their renaturation, at least partially. The yield depends on the components of the diluent and the results shown in Table 1 could probably be improved further. The effect of the components is unpredictable; for example, use of Triton X-100 or other detergents generally increased the yields, but reduced that of protein # 4.

The resulting aqueous and 75% glycerol solutions are not always stable, and further precipitation can take place. In the case of proteins # 1, solutions as concentrated as 10 mg protein/ml could be obtained in the final dilution step. However, these slowly formed precipitates over the course of several days at 4°C, stabilizing only when the concentration reached 2 mg/ml. Inclusion of Triton X100 prevented this instability. Proteins # 3 precipitated in the 75% glycerol solution after a storage of three months at −20°C.

The exact basis of the improvement obtained by this procedure is not known. We can speculate that the increase in viscosity might prevent aggregation and precipitation, and that the high H-bond forming character of glycerol might stabilize protein conformations. Indeed, proteins seem to interact with glycerol similarly to water, because lysozyme keeps its native conformation in glycerol (Ref. 6). This procedure has been under patent application (Ref. 7).

Acknowledgements

The work was supported by an International Collaboration Grant of Japanese Ministry of Education, Science and Culture to NS (09044238).

References

[1] Mukhopadhyay A. (1997)
Adv. Biochem. Eng. Biotechnol. , **56**:61-109. Cited by

[2] Rudolph R. and Lilie H. (1996)
FASEB J. , **10**:49-56. Cited by

[3] Grisshammer, R. and Nagai, K. (1995) in *DNA Cloning 2: Expression Systems: A Practical Approach* (Glover, D.M. and Hames, B.D., eds), pp. 59–92, IRL Press

[4] Fernandez-Patron C., Castellanos-Serra L. and Rodriguez P. (1992)
BioTechniques , **12**:564-573. Cited by

[5] Ortiz M.L. *et al.* (1992)
FEBS Lett. , **296**:300-304. Cited by

[6] Rariy R.V. and Klibanov A.M. (1997)
Proc. Natl. Acad. Sci. U. S. A. , **94**:13520-13523. Full text Cited by

[7] Japan. Pat. Int. Cat. B01J 19/00 (#144758)

[8] Nagai H. and Shimamoto N. (1998)
Gene. Cells , **2**:725-734.

Glossary

Products Used

protein assay
protein assay from Bio-Rad
DEAE Toyopearl S650 resin
DEAE Toyopearl S650 resin from Toyobo

PII: S0168952598015765

Copyright